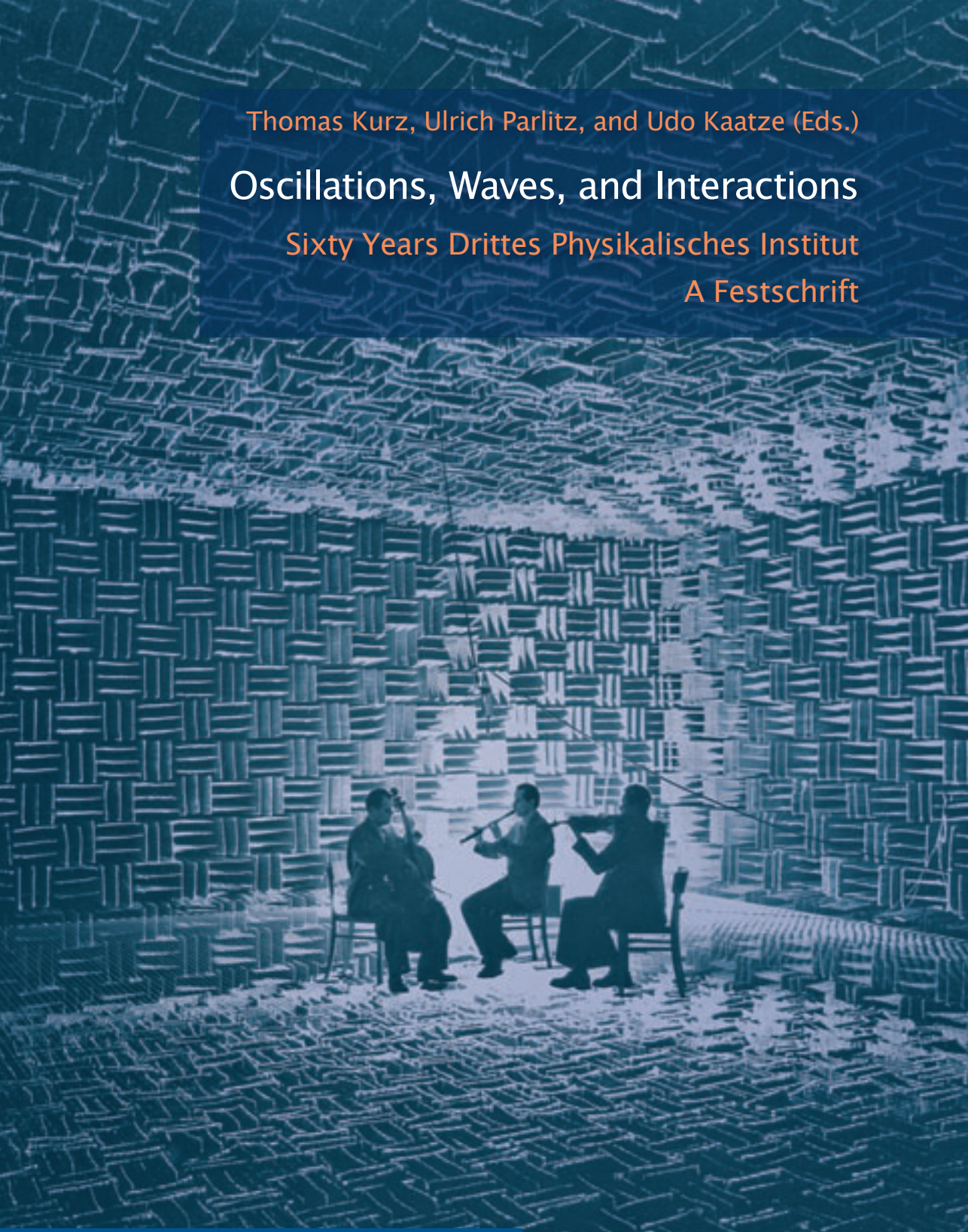
Thomas Kurz, Ulrich Parlitz, and Udo Kaatze (Eds.)

Oscillations, Waves, and Interactions

Sixty Years Drittes Physikalisches Institut

A Festschrift



Universitätsverlag Göttingen

Thomas Kurz, Ulrich Parlitz, and Udo Kaatze (Eds.)

Oscillations, Waves, and Interactions

erschienen im Universitätsverlag Göttingen 2007

Thomas Kurz, Ulrich Parlitz,
Udo Kaatze (Eds.)

Oscillations, Waves and Interactions

Sixty Years Drittes Physikalisches Institut

A Festschrift

Universitätsverlag Göttingen
2007

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar

Address of the Editors
Drittes Physikalisches Institut,
Friedrich-Hund-Platz 1
D-37077 Göttingen
office@physik3.gwdg.de; http://www.physik3.gwdg.de/

Typesetting in LaTeX: Thomas Kurz
Cover Design: Margo Bargheer
Cover image: Institute's free-space room. A cut-out of this image appeared in the LIFE International special issue "Germany, a giant awakened", May 10, 1954.

http://univerlag.uni-goettingen.de
ISBN: 978-3-938616-96-3

Contents

Preface

This year we celebrate the sixtieth anniversary of the Dritte Physikalische Institut. Professor Erwin Meyer, the first director, soon after the foundation formed the institute's own style. He established various groups which, headed by a senior scientist, dealt with quite different fields of research. Based on an atmosphere of academic freedom and confidence these groups did not just complement one another. Mutual support with equipment and financial assistance among the groups went without saying.

This spirit has been perpetuated by Erwin Meyers former students, in particular by his successors Manfred R. Schroeder and Werner Lauterborn, but also by Reinhard Pottel, Hans-Wilhelm Helberg, Dirk Ronneberger, and Dieter Guicking, who spent their academic career in the institute. It has promoted a broad spectrum of research topics which were and still are conjointly treated, bond by the concerted theme "oscillations and waves".

In this book some review articles by former and present members of the Dritte are combined, indicating the broadness of the research interests which developed from this theme. Unfortunately, such collection of articles allows only for a limited access to the broad variety of different fields of interest that have been dealt with in the past. The reviews are, therefore, complemented by a short overview which, again, cannot show all aspects of successful work during the sixty years of existence of the institute. Also, a short outlook is given at the end of the book on future research objectives in complex systems and biophysics.

We would like to thank all colleagues who established by their activity to the reputation of the institute and who thus rendered this book possible. We thank the technicians who by their skilful constructions of new apparatus enabled many new and more precise methods of measurements, and we thank the tracer, photographer, and administration secretaries for their continual support. We are particularly indebted to the authors who spent much time to contribute to this book.

Göttingen, September 2007

Thomas Kurz
Ulrich Parlitz
Udo Kaatze

Oscillations, Waves and Interactions, pp. 1–24
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-01-8

Applied physics at the "Dritte" Fruitful interplay of a wide range of interests

Manfred R. Schroeder, Dieter Guicking, and Udo Kaatze

Drittes Physikalisches Institut, Georg-August-Universität Göttingen,
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

1 Historical antecedents

At the end of the 19th century the well-known Göttingen mathematician Felix Klein proposed to promote the pursuit of applied sciences at German universities. He was encouraged in this effort by a visit to the 1893 Chicago World Exhibition – as an official representative of Kaiser Wilhelm II – and subsequent visits to several American universities, which had a strong tradition of fostering applied sciences and engineering. But Klein's attempt to enlist the Technische Hochschule Hannover – let alone any university – in his endeavour failed miserably. In 1898 he succeeded, with support by German industry (Böttinger), to establish an Institut für Angewandte Elektrizität and an Institut für Angewandte Mathematik und Mechanik at Göttingen,

Institut für Angewandte Elektrizität – the "Red House" on Bunsenstraße.

The institute in the 1950s – the "White House" on Bürgerstraße, the former Institut für Angewandte Mathematik und Mechanik.

the latter became home to such famous aero-dynamicists as Ludwig Prandtl and Theodor von Kármán.

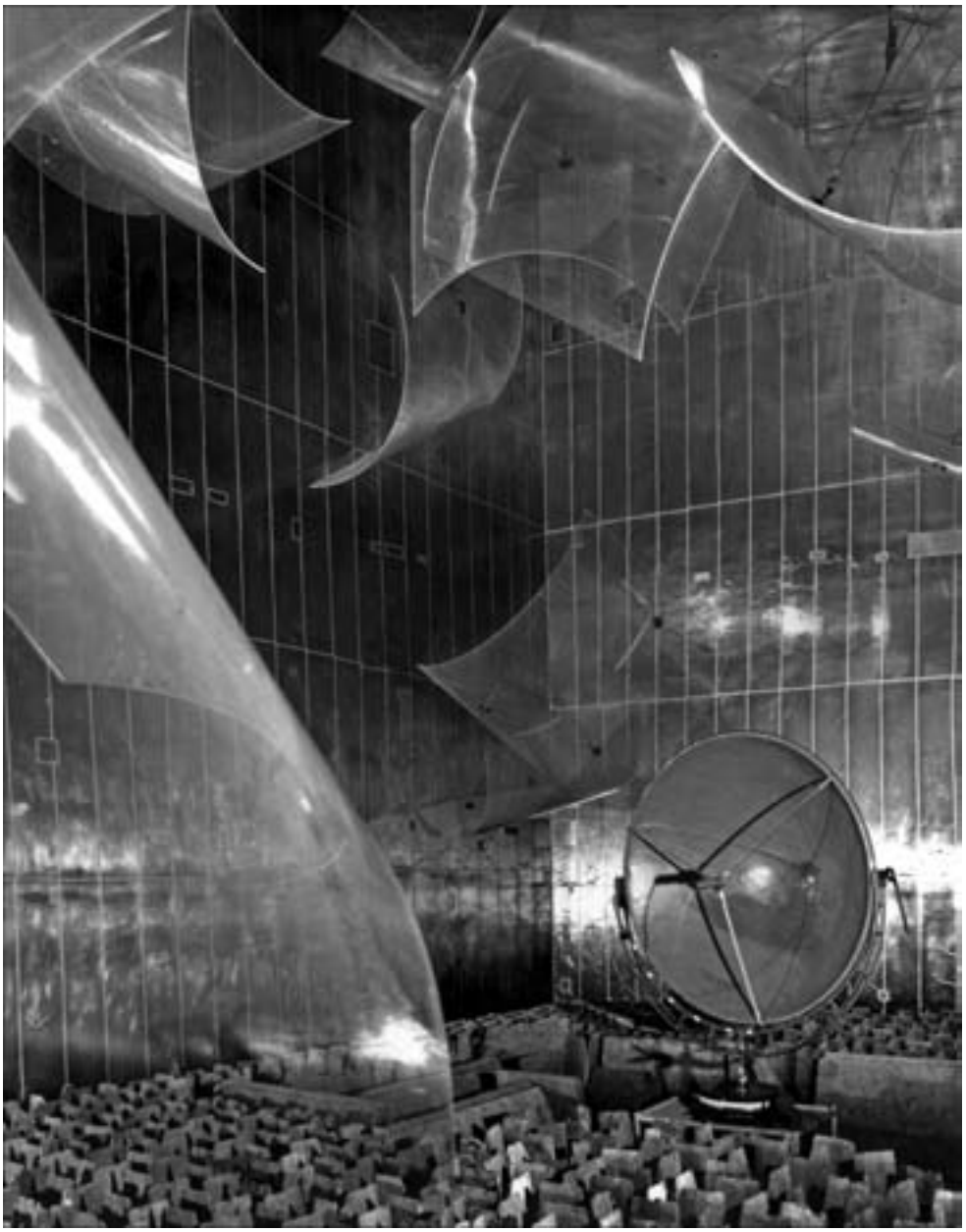
In May 1947 these institutes morphed into the Dritte Physikalische Institut under the leadership of Erwin Meyer, who had been professor at the Technische Hochschule Berlin-Charlottenburg and, at the same time, head of the acoustics and mechanics department at the Heinrich-Hertz-Institut, Berlin. With Meyer, room acoustics, including concert hall acoustics, underwater sound, and acoustics in general, as well as microwaves came to Göttingen.

2 Room acoustics

Göttingen distinguished itself not only in fundamental research in room acoustics, but also by practical applications and acoustic consulting. The Institut became soon well-known through several successful large projects: the Beethovenhalle in Bonn [1], the Liederhalle [2] and the plenary hall of the Baden-Württemberg parliament in Stuttgart [3], and the Jahrhunderthalle in Höchst [4]. Especially noteworthy was the ingenious utilisation of electro-acoustic means utilising the Haas-Effekt.

In fact, the Institut became well known internationally by the discovery, stimulated by Meyer, of the Haas-Effekt by Helmut Haas in 1951 [5]. This property of human auditory perception says that even an amplified – but suitably delayed – sound will not affect the perceived direction of the sound source. The name Haas effect was suggested by Richard Bolt of the Massachusetts Institute of Technology. It soon became the basis of electro-acoustic installations using "Leisesprecheranlagen" (gentle-speaker facilities) [6].

“Free-space” room with Frieder Eggers (cello), Wolfgang Westphal (flute) and Heiner Kuttruff (violin).



Reverberation room.

Sound diffusion in small and large rooms, radio studios as well as concert halls, was investigated by Rolf Thiele, whose results were made visible by a "Schalligel" ("sound hedgehog") giving the different sound intensities in different spatial directions [7]. Meyer and Thiele also introduced the concept of "Deutlichkeit" (definition), the sound energy in the first 50 milliseconds of the impulse response [8].

In general, there was a world-wide interest to supplement reverberation time by other physical parameters to characterise the acoustic quality of performance spaces. One of these criteria, favoured in the early 1950s, was the so-called "frequency-irregularity" of the sound transmission between source (on the stage) and a listener's ears. Experimental results obtained by Heinrich Kuttruff and Rolf Thiele [9] in the Herkulessaal in Munich (and other concert halls) showed that the number of maxima of the frequency response did not correspond to the number of normal modes (resonances) – as posited by a faulty theory by Bolt – but was actually more than a thousand times less. This astonishing result was explained by a statistical theory by Manfred Schroeder, then a post-doc at the institute [10]. This theory showed that the frequency response (sound pressure and phase as functions of frequency) is, approximately, a complex Gaussian process. As a result, the average distance between maxima above a critical frequency ("Schroeder frequency") is fully determined by the (reciprocal) reverberation time – thus not affording the much sought-after new quality parameter.

Beside measurements in actual rooms, scale models were also studied. Following Meyer's inclination to consider sound waves and electromagnetic waves simultaneously, microwave models were also included. Schroeder could show that the distribution of the frequencies and excitations of the normal electromagnetic modes in metallic cavities were highly irregular – even for very small deviations from the symmetry of a perfectly rectangular space, such as a cube [11]. Thus, for all practical purposes in room acoustics, the normal modes (resonances) can be considered completely random.

Fundamental investigations, such as on the perception of "echoes", were also conducted. For this purpose, the attic of the "white house" on Bürgerstraße was converted into a makeshift "anechoic" space. Later a large Reflexionsfreier Raum ("free-space room") was constructed, which – unique in the world – was also designed to be nearly free from reflections for microwaves to facilitate free-space measurements with electromagnetic waves [12]. Also a large reverberation room ("Hallraum") was built which, again, was reverberant for both sound waves and electromagnetic waves [13].

In 1963, Meyer and Kuttruff studied the reflective properties of the ceiling of Philharmonic Hall in New York by means of a scale model leading to an explanation of the observed low-frequency deficiency in the actual hall [14]. In the early 1970s, in a large study of concert hall quality, Dieter Gottlob, Manfred Schroeder, and Karl-Friedrich Siebrasse – on the basis of measurements in 22 concert halls in Europe – showed that the lack of early lateral reflections in many modern halls with low ceilings and wide (fan-shaped) ground plan was the main culprit [15]. To counteract this deficiency, Schroeder conceived, after 1975, sound-diffusing structures ("reflection phase gratings") based on number-theoretic principles [16], which have found broad acceptance in room acoustics. At the same time, Schroeder proposed a new method of accurately measuring reverberation times by "backward integration" of the (squared)

impulse response of a room using so-called maximum-length sequences constructed from the theory of finite number fields [17].

The room-acoustic tradition of the institute lives on in the acoustic consultancies of former students (Akustikbüro Göttingen).

3 Speech and hearing

When Manfred Schroeder, who had worked at Bell Telephone Laboratories since 1954, succeeded Erwin Meyer as director, some of the speech research pursued at Bell was transplanted to Göttingen in 1969. Much of the speech research at Göttingen involved the application of measuring methods from physics to the human speaking process. The work included speech synthesis [18,19], prosody [20], speech and speaker recognition [21,22], speaker-specific vocal-tract parameters [23] and used advanced mathematical methods, such as neural networks and hidden-Markov processes.

Impedance measurement of the lips and the glottis [24] were performed. Another goal was the deduction of the area function and of articulatory parameters from acoustic data, such as the speech signal [25,26], from lip photography and X-rays.

The cross-fertilisation between speech and hearing research is exemplified by the investigation of the cocktail-party processors by H. W. Strube et al. [27]. Also modulation-frequency filtering was applied [28]. For fundamental reasons, and in view of the importance of human hearing for the proper encoding of speech and music signals, extensive studies of perceptual masking of one sound by another were undertaken [29]. This research was led by Birger Kollmeier and Armin Kohlrausch. A considerable amount of the work was concerned with the design of better hearing aids [30] and tests of speech intelligibility for hearing-impaired listeners.

The work on speech compression was based on linear predictive coding (LPC). In 1976, Atal, Hall and Schroeder introduced "perceptual coding" to acoustic signals resulting in high-quality speech at very low bit-rates, essential for cell phones and Internet applications.

Some of the ongoing work in speech is aimed at improving diagnostic tools for voice pathologies (in collaboration with professor E. Kruse, see the contribution in this book [31]). Hearing research is still a field of interest of B. Kollmeier at the University of Oldenburg and A. Kohlrausch at Philips Research, Eindhoven [32].

4 Noise control

In the early 1980s, research projects on active impedance control were started with controllers in analog electronics, both for air-borne sound and structural vibrations [33,34]. Besides several smaller projects, active broadband noise control in cars was investigated in collaboration with an automobile manufacturer, applying adaptive digital feedforward control with fast algorithms to cope with nonstationary tyre rolling noise, and with varying acoustic transfer functions [35]. As a demonstration object, low-frequency fan noise of a kitchen exhaust was cancelled successfully [36]; a presentation at the Hanover fair 1995 found vivid interest. A major research field

was the improvement of adaptive filter algorithms with fast convergence, yet low numerical complexity [37,38].

These activities ended with the retirement of Dieter Guicking (1998); some more projects on active flow control and related problems were performed in the working group of Dirk Ronneberger, see Section 5 and his contribution to this book [39]. For more details on the historical development, fundamentals, and the state of the art of active sound and vibration control see the article by D. Guicking [40] in this book.

5 Flow acoustics and flow control

Sound propagation through ducts is influenced by superimposed fluid flow. These interactions have been studied at the institute theoretically and experimentally in research groups headed by Fridolin Mechel (1960–1966) and thereafter by Dirk Ronneberger until to his retirement 2006. Experimental facilities comprised wind tunnels and ducts with recirculating water or oil flow, the latter because of the thicker boundary layer. In the wind tunnels, the interaction of sound propagation and mean flow was studied under various aspects: in ducts with absorbing lining without and with absorber cassetting [41], in rigid-walled ducts with turbulent air flow [42], also under the influence of wall roughness [43]. Much work was devoted to the investigation of cross-sectional discontinuities [44], acoustical [45] and vibrational [46] control of the boundary layer, the acoustic impedance of an orifice in the flow duct [47] or at the side wall [48], the directivity of sound radiation from the duct end [49], and the influence of nonlinear interaction of instability waves in a turbulent jet [50].

In the 1980s, intensive experiments were performed on the noise radiation of rolling automobile tyres and have contributed to a better understanding of the noise generation mechanisms [51].

Since the late 1990s, the focus was shifted to active control of flow parameters, both in air and water [52–56]. Some recent research results are presented at another place in this book [39].

6 Underwater sound

Erwin Meyer and his team had done intensive research work on water-borne sound until 1945 at the Heinrich-Hertz-Institut in Berlin. They studied the mechanisms of sound absorption in sea water, and they developed new absorbers of underwater sound: rib-type absorbers for the lining of anechoic water basins, and thin-layer two-circuit resonance absorbers as reflection reducing coatings for underwater objects. The research results, which could not be published during war time, have been collected in a US Navy Report [57] which today is still recognised as a treasure of information for researchers in this field all over the world.

Meyer and his team continued the hydroacoustic research at Göttingen with financial support by the British "Department of Scientific and Industrial Research" (DSIR), later by the "Department of Naval Physical Research" (DNPR). The research contract started in 1948 and was continued year after year for a record breaking period of time until 1978, long after Erwin Meyer's death.

Anechoic water tank lined with rip-type absorbers.

Sound absorption and sound velocity in water, aqueous solutions of electrolytes, and other liquids were studied in wide frequency and temperature ranges (see Section 10). Also, the influence of gas bubbles on the acoustic properties of liquids was investigated [58,59], further leading to intensive research on single bubble oscillations, cavitation, and luminescence (see Section 7 of this article and the pertinent contributions to this book [60,61]).

Erwin Meyer was always interested in the improvement of measurement techniques, also in the field of underwater sound. In close analogy to the anechoic chambers for air-borne sound, for which Meyer had invented the lining by wedges of sound absorbing porous material [62], previous experiments [63] eventually culminated in the installation of a 100-m^3 water tank with rib-type absorbers [64], providing near free-field conditions from 7 kHz through 70 kHz. Less costly impedance measurements in resiliently lined water-filled tubes [65] and sound absorption or sound power measurements in a resiliently lined reverberation tank [66] were also performed. As an interesting result it was shown that the edge effect is much stronger for absorbers on sound-soft than on rigid walls [67]. Impedance measurements at low frequencies were performed with thick-walled "pressure chambers" [68]. A shallow water basin with sound-soft bottom and absorbing perimeter allowed studies of two-dimensional sound fields [69].

The scattering of underwater sound was investigated in our institute for more than 30 years, starting 1965 with sound-soft objects [70], and since 1989 resonance scattering with possible application to the detection and classification of objects buried in marine sediments [71]. Motivated by analogous microwave systems, Peter Wille constructed a slim directional hydrophone with low flow resistance (in analogy to the microwave "rod radiator") [72], and an aspect-independent sonar reflector, based on inhomogeneous lenses [73].

Upon Erwin Meyer's initiative, the international Journal *Acustica* was founded in Rome 1950, to succeed the German *Akustische Zeitschrift* (1936–1944). Since much unpublished work had accumulated in the meantime, the editors of Acustica published *Akustische Beihefte* as supplements in which these papers were collected, presenting, among others, the major results of the war-time work of Meyer's group on hydroacoustics. One such paper [74] describes the thin-layer two-circuit resonance absorbers – combinations of a parallel and a series resonance circuit – originally made of rubber with air holes in a central layer. With the availability of a great variety of modern high polymers, it was attempted to improve the resonance absorbers by searching for better suited materials. Since a material with a high loss factor of the bulk modulus could not be found, polymers with lossy Young's and shear moduli were applied [75,76]. It was, however, soon recognised that the desired high loss factor was inevitably related to a strong frequency and temperature dispersion of the modulus because they are coupled by the Kramers-Kronig relations (e. g., [77]), which made practical applications unrealistic. As an alternative, the absorption of sound by constrained liquid flow was investigated, and various structures were designed [78,79]. With electrorheological fluids (ERF), absorbers have been developed the properties of which can be electrically adjusted to account for a changing environment such as hydrostatic pressure (i. e., depth of water) or temperature [80].

Inspired by the success of these absorbers, a two-circuit resonance absorber of

Drittes Physikalisches Institut at the turn of the millenium.

airborne sound was constructed for its potential to attenuate the rolling noise of tyres, covering the relevant frequency range from 700 Hz through 1300 Hz [81].

Some basic research was done on the question "How does the sound get out of a ship into the water". Experiments and analytical models of sound radiation from flat plates [82] were followed by investigations with thick-walled cylindrical shells. The calculation of their resonance frequencies (at which the sound radiation is maximum) is all but trivial – the rigorous theory demands for the solution of partial differential equations of 10th order [83]. Attempts to find simpler calculations resulted in a surprisingly good approximation [84], reducing the computational effort to solving a third-order algebraic equation. Comparison with experimental data revealed errors of a few percent only. As an aside, the critical frequency of flat plates turned out to be strongly dependent on the density of the surrounding medium [85].

The activities on underwater sound research, except for cavitation, ended with the retirement of Dieter Guicking who has supervised much of the work summarised here since about 1970.

7 Cavitation

From the very beginning cavitation was among the areas of research of the Dritte Physikalische Institut. In fact, already in Berlin Meyer and Tamm had been concerned with vibrations of bubbles in liquids [86]. Methods to study cavitation phe-

nomena using ultrasound were immediately developed in Göttingen [87]. A cinematograph was constructed that allowed for picture sequences of oscillating bubbles at a picture repetition rate up to 65 000 per second [88]. Soon interest was directed towards nonlinear vibration characteristics of bubbles in water [89] and, in a series of papers [90–92], the occurrence of sonoluminescence in the sonically induced bubbles was reported.

Lauterborn generated cavitation bubbles by focusing giant pulses from a ruby laser into liquids [93]. He investigated the bubble dynamics by high-speed cinematography with picture repetition rates up to 900 000 per second [94]. For his fundamental contributions to the field he was awarded the Physikpreis of the Deutsche Physikalische Gesellschaft (German Physical Society) in 1976.

Since then many aspects of cavitation are being investigated by Lauterborn and his group. Bubble dynamics was studied experimentally and theoretically [60,61, 95]. Much emphasis was placed on the spatial distribution of bubbles in cavitation structures and on the collective phenomena of bubble clusters. The chaotic behaviour of bubbles as reflected in the properties of sound waves radiated from cavitation structures opened new vistas in nonlinear acoustics of fluids [96] and added a new experimental branch to nonlinear physics [97,98]. It had a considerable influence on various of the institute's areas of research [99] (see section 9 of this article). The fascinating features of sonoluminescence have been an enduring topic of interest [100, 101]. Also neutrons in close temporal proximity to cavitation luminescence were searched for [102].

An appealing characteristic of the field is the combination of basic research with a multitude of applications. Biologically and medically related applications are briefly mentioned in section 13 on "Biophysics". Some are described in greater detail in other contributions to this book [103,104]. Use of high power ultrasound in chemical processing and cleaning is also based on energy concentration by bubble collapse [105, 106].

8 Optics and Holography

Common photographic techniques and high-speed cinematography are rather limited when three-dimensional bubble clusters are to be recorded. Superior images of cavitation structures were obtained using holography [107]. Shortly after the invention of the laser, this method to store three-dimensional images became very popular worldwide. At the institute holography was not just used for imaging but also for generation of cavitation bubble systems [108].

Werner Lauterborn and Karl Joachim Ebeling took a big step forward by conceiving high-speed holocinematography [109,110]. Spatial and spatial-frequency multiplexing techniques were applied to achieve image separation during the reconstruction process. Acousto-optic modulators [110,111] were employed as beam splitters and deflectors. Up to eight holograms were successively recorded at a maximum frame rate $20\,000\,\mathrm{s}^{-1}$ to show the interactions of laser-produced bubbles. Combining rotation of the holographic plate and acousto-optic beam deflection, the capability of high-speed holocinematography was extended to record up to 4,000 holograms

of rapidly changing deep objects with frame rates up to $300\,000\,s^{-1}$, corresponding to a 3-μs time interval between two pictures [112,113]. Such advanced holographic methods were used to investigate time-dependent cavitation phenomena, such as period-doubled and chaotic bubble oscillations [114].

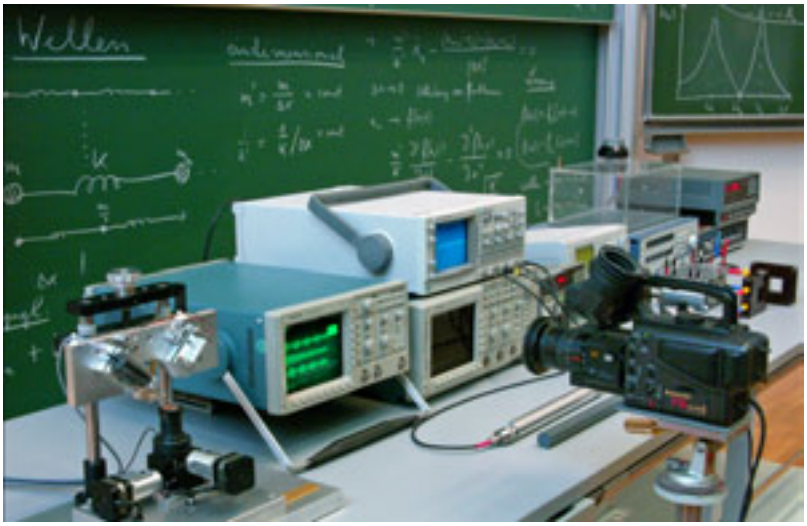
9 Nonlinear dynamics

Much of classical physics is based on linear laws. But in recent decades nonlinearity has come to the fore. With Werner Lauterborn and his coworkers nonlinear dynamics has been added to the research repertoire at the Dritte Physikalische Institut [98,115]. In particular Lauterborn studied the nonlinear oscillations of cavitation gas bubbles in liquids which show nonlinear resonances [116], period-doubling bifurcations, and chaotic dynamics [117,118]. Soon thereafter the Lauterborn group investigated chaotic dynamics and bifurcations in periodically excited nonlinear oscillators [119–121] and coupled oscillators [122,123]. Later Lauterborn and others shed light on the nonlinear dynamics of lasers [124–126] and nonlinear waves and solitons [127,128]. In seperate investigations, Parlitz and colleagues studied the synchronisation properties of chaotic systems, methods for time series analyses and predictions, and also control of chaos. Part of these investigations is described in more detail in the article by Ulrich Parlitz [99] of this book.

10 Complex liquids

In the years before the invention of the laser in 1960, microwaves were used in demonstration experiments [129] requiring coherent electromagnetic signals, and in diffraction studies [130,131]. At that time much work was devoted to the development of microwave techniques, including transmission lines [132], antennas [133], and absorbers [134–137]. This applied research soon induced interest in the principles of molecular systems. Electromagnetic waves were used to investigate aspects of ferroelectricity [132,138–141] and ferromagnetism [142,143] and to perform dielectric studies of the molecular behaviour of liquids [144]. The application to molecular physics was greatly facilitated by an extensive collection of radio frequency and microwave devices at the institute and, thanks to the excellent support by the electronics and precision engineering workshops, the Dritte became soon well known for its sophisticated broadband measuring methods in liquid research. Later ultrasonic attenuation spectrometry was extended to cover a very broad frequency range [145–147]. Additional methods, such as shear wave spectrometry [148] and dynamic light scattering [149] were also used.

Peter Debye, professor for theoretical and experimental physics at Göttingen from 1914 to 1920, was the first to illuminate the molecular aspects of the interactions of electromagnetic waves with materials. A significant step towards dielectric spectroscopy of liquids was Reinhard Pottel's broadband study of 2:2 valent electrolyte solutions in which he verified the existence of dipolar ion complex structures [150], as had been suggested by Günter Kurtze, Konrad Tamm and Manfred Eigen on the basis of ultrasonic spectroscopy [146,151,152].



Lectures in all fields of vibrations and waves, usually supplemented by experiments, have been a continual element of academic life at the “Dritte”.

Pottel’s idea lead to more than forty years of research into the molecular behaviour of liquids [146,147]. Although other aspects, such as the dielectric [153,154] and ultrasonic [155] relaxations of non-dipolar liquids, were also considered, attention was predominantly directed towards hydrogen network fluctuations, self-associations, and conformational variations in associating systems. Because of its omnipresence on our planet, special attention was given to water in its different states of interaction. More recently, critical phenomena and their crossover to the noncritical dynamics of demixing binary and ternary liquids were (and still are) the focus of interest. Some aspects are summarised in contributions to this book [146,147].

11 Hypersonic spectroscopy of solids

The discovery of surface excitation and detection by Baransky and by Bömmel and Dransfeld in the late 1950s opened up new vistas for high-frequency ultrasonic investigations of materials. Klaus Gottfried Plaß took the opportunity to develop an ultrasonic spectrometer for liquid attenuation measurements up to GHz frequencies [145]. Wolfgang Eisenmenger and his group performed a comprehensive investigation of the hypersonic attenuation in cylindrical quartz rods up to 10 GHz [156]. Measurements at temperatures between 4 K and 273 K involved a helium liquefier which subsequently was used in various low temperature investigations.

A breakthrough in phonon spectroscopy was reached by Eisenmenger during a stay at Bell Telephone Laboratories in the summer 1966. Together with A. H. Dayem he succeeded in quantum generation and detection of incoherent phonons, using superconducting tunnel diodes [157,158]. Hypersonic measurements at frequencies between 1 GHz and 1 THz were made possible by this pioneering work.

12 Organic conductors and deformed semiconductors

The transition from more applied research in radio-frequency and microwave techniques to applications of such techniques in basic research also happened by the orientation towards organic conductors. Interest in these materials springs from their molecular structure which promotes electrical conductivity in one direction. This confinement to quasi-one-dimensional behaviour opens insights in interesting phenomena which do not exist or can hardly be studied in three dimensions [159]. Additionally, one-dimensional conductors are considered potential materials for applications.

Hans-Wilhelm Helberg and his group investigated a multitude of organic conductors, semiconductors, and superconductors, comprising polymers, charge transfer complexes, and doped materials [160–164]. DC and microwave conductivity as well as electric permittivity measurements were performed in wide ranges of frequency (0.5 to 60 GHz) and temperature (1.7 to 700 K) in order to elucidate the mechanisms of electronic excitation and conductivity. Intra- and intermolecular excitations have been considered using polarisation microscope techniques in the visible and near infrared spectral ranges to determine the indicatrix orientation and the directions of optical absorption. One-dimensional solids are further on studied by Martin Dressel, a former student of the Helberg group, at the University of Stuttgart [159].

13 Biophysics

Biologically inspired or medically motivated topics were considered by several groups of the institute. Molecular aspects apply to head group reorientations of zwitterions, hydrocarbon chain isomerisations, domain structure fluctuations and defect formations in lipid membranes [165–167] as well as to the molecular dynamics and conformational kinetics of carbohydrates in solution [168,169]. Some of these investigations benefited much from close cooperation with colleagues at the Max-Planck-Institut für Biophysikalische Chemie [170–174].

A quite different approach to biophysics was the treatment of nonlinear differential equations to gain a better understanding of the biological clock [175]. The numerical solution of the homogeneous and inhomogeneous van der Pol equation was aimed at entrainment-synchronization phenomena of a self-excited oscillator by an external one.

Increasing demands in the monitoring of organ properties during ischemia sparked interest in the non-invasive dielectric spectroscopy of organ tissue [176] and in the modelling of the electrical impedance of cellular media [177].

In the 1980s use of focused pulses from a neodynium:YAG laser enabled non-invasive intraocular surgery by photodisruption. This new instrument called for a careful investigation of acoustical transients and cavitation bubbles by laser-induced breakdown in liquids on conditions relevant to ophthalmologic applications [178,179]. Lithotripter shock waves cause cavitation inside the human body. The action of cavitation bubbles leads to tissue damage as side effect of narrow-focus extracorporal shock wave lithotripsy [103]. Shock waves and ultrasound may also enhance the transient permeability of cell membranes and may facilitate drug delivery without

permanent damage to the cell, a process called sonoporation. Molecular uptake is a precondition for many biological and medical applications. In order to gain deeper insights in both mechanisms, tissue injury and sonoporation, the interactions of cavitation bubbles with cells and boundaries have been studied [180,181].

The need for an objective assessment of voice quality after curative minimal-invasive laser resection of laryngeal carcinomas led to an intensive cooperation with the department of phoniatrics and pedriatric audiology of our university [182,183]. Acoustical analyses for the description of pathologic voices and for the documentation of progress in voice rehabilitation measures were developed [31].

In 2000 a biophysics group was established at the Dritte. Headed by Manfred Radmacher this group was concerned with mechanical properties of biomaterial and imaging of biological samples using atomic force microscopy [184,185].

In 2006 Christoph Schmidt was appointed professor for biophysics at the institute. The general theme relating his research projects is the study of dynamic physical properties of complex biological macromolecules and their assemblies up to the level of cells [186–188]. These projects involve a) biological motor proteins in single-molecule experiments with the goal of understanding the physical principles of biological force generation in a multitude of active transport processes, b) the dynamics of DNA enzymes in order to understand the still incompletely known highly complex mechanical tasks in replication, transcription and packing, and c) the collective dynamics of systems ranging from synthetic colloids to whole cells. In particular, networks of semiflexible proteins are studied, in vitro and also in living cells, with so-called microrheology techniques in order to understand the functional principles of the cytoskeleton, which play crucial roles in processes such as cell division, cell locomotion, or cell growth, as well as mechano-sensing and signalling. Future work is indicated in the article by Christoph Schmidt and Stefan Lakämper in this book [189].

References

[1] E. Meyer and K. H. Kuttruff, 'Zur akustischen Gestaltung der neuerbauten Beethovenhalle in Bonn', Acustica **9**, 465 (1959).
[2] W. Junius, 'Raumakustische Untersuchungen mit neueren Meßverfahren in der Liederhalle Stuttgart', Acustica **9**, 289 (1959).
[3] E. Meyer and K. H. Kuttruff, 'Die raumakustischen Maßnahmen beim Neubau des Plenarsaals des Baden-Württembergischen Landtages in Stuttgart', Acustica **12**, 55 (1962).
[4] E. Meyer and K. H. Kuttruff, 'Zur Raumakustik einer großen Festhalle (Erfahrungen mit einer elektroakustischen Nachhallanlage)', Acustica **14**, 138 (1964).
[5] H. Haas, 'Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache', Acustica **1**, 49 (1951).
[6] G. R. Schodder, F. K. Schröder, and R. Thiele, 'Verbesserung der Hörsamkeit eines Theaters durch eine schallverzögernde Leisesprechanlage', Acustica **2**, 115 (1952).
[7] E. Meyer and R. Thiele, 'Raumakustische Untersuchungen in zahlreichen Konzertsälen und Rundfunkstudios unter Anwendung neuerer Meßverfahren', Acustica **6**, 425 (1956).
[8] E. Meyer, 'Definition and Diffusion in Rooms', J. Acoust. Soc. Am. **26**, 630 (1954).

[9] K. H. Kuttruff and R. Thiele, 'Über die Frequenzabhängigkeit des Schalldrucks in Räumen', Acustica **4**, 614 (1954).
[10] M. R. Schroeder, 'Die statistischen Parameter der Frequenzkurven von großen Räumen', Acustica **4**, 594 (1954).
[11] M. R. Schroeder, 'Eigenfrequenzstatistik und Anregungsstatistik in Räumen. Modellversuche mit elektrischen Wellen', Acustica **4**, 456 (1954).
[12] E. Meyer, G. Kurtze, H. Severin, and K. Tamm, 'Ein neuer großer reflexionsfreier Raum für Schallwellen und kurze elektromagnetische Wellen', Acustica **3**, 409 (1953).
[13] E. Meyer, G. Kurtze, K. H. Kuttruff, and K. Tamm, 'Ein neuer Hallraum für Schallwellen und elektromagnetische Wellen', Acustica **10**, 253 (1960).
[14] E. Meyer and K. H. Kuttruff, 'Reflexionseigenschaften durchbrochener Decken (Modelluntersuchungen an der Reflektoranordnung der neuen Philharmonic Hall in New York)', Acustica **13**, 183 (1963).
[15] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, 'Comparative Study of European Concert Halls: Correlation of Subjective Preference with Geometric and Acoustic Parameters', J. Acoust. Soc. Am. **56**, 1195 (1974).
[16] M. R. Schroeder, *Number Theory in Science and Communication – With Applications in Cryptography, Physics, Digital Information, Computing and Self-Similarity*, Springer Series in Information Sciences 7 (Springer, Berlin, 2006), 4th ed.
[17] K. H. Kuttruff and M. J. Jusofie, 'Nachhallmessungen nach dem Verfahren der integrierten Impulsantwort', Acustica **19**, 56 (1967).
[18] H. W. Strube and R. Wilhelms, 'Synthesis of unrestricted German speech from interpolated log-area-ratio coded transitions', Speech Communication **1**, 93 (1982).
[19] P. Meyer, R. Wilhelms, and H. W. Strube, 'A quasiarticulatory speech synthesizer for German language running in real time', J. Acoust. Soc. Am. **86**, 523 (1989).
[20] Z. Antoniadis, *Grundfrequenzverläufe deutscher Sätze: Empirische Untersuchungen und Synthesemöglichkeiten*, Dissertation, Universität Göttingen (1984).
[21] H. W. Strube, D. Helling, A. Krause, and M. R. Schroeder, 'Word and Speaker Recognition Based on Entire Words without Framewise Analysis', in *Speech and Speaker Recognition*, edited by M. R. Schroeder (Karger, Basel, 1985), pp. 80–114.
[22] T. Gramß and H. W. Strube, 'Recognition of isolated words based on psychoacoustics and neurobiology', Speech Communication **9**, 35 (1990).
[23] H. Freienstein, K. Müller, and H. W. Strube, 'Vocal-tract parameter estimation from formant patterns', Acustica **85**, 52 (1999), Joint Meeting ASA/EAA/DEGA, Berlin 1999. Full paper on CD-ROM only.
[24] S. Rösler and H. W. Strube, 'Measurement of the glottal impedance with a mechanical model', J. Acoust. Soc. Am. **86**, 1708 (1989).
[25] M. R. Schroeder and H. W. Strube, 'Acoustic Measurements of Articulator Motions', Phonetica **36**, 302 (1979).
[26] G. Panagos, *Messung artikulatorischer parameter und Schätzung des artikulatorisch-akustischen Zusammenhangs*, Dissertation, Universität Göttingen (1989).
[27] H. W. Strube, 'Separation of several speakers recorded by two microphones (cocktail-party processing)', Signal Processing **3**, 355 (1981).
[28] T. Langhans and H. W. Strube, 'Speech enhancement by nonlinear multiband envelope filtering', in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP-82)* (IEEE, New York, 1982), paper S1.3, pp. 156–159.
[29] A. Kohlrausch, 'Auditory Filter Shape Derived from Binaural Masking Experiments', J. Acoust. Soc. Am. **84**, 573 (1988).
[30] B. Kollmeier, 'Speech Enhancement by Filtering in the Loudness Domain', Acta Otolaryngol. Suppl. **469**, 207 (1990).

[31] H. W. Strube, ‘Speech research with physical methods’, in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[32] A. Kohlrausch and S. van de Par, ‘On the use of specific signal types in hearing research’, in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[33] D. Guicking, K. Karcher, and M. Rollwage, ‘Coherent active methods for applications in room acoustics’, J. Acoust. Soc. Am. **78**, 1426 (1985).
[34] D. Guicking, J. Melcher, and R. Wimmel, ‘Active Impedance Control in Mechanical Structures’, Acustica **69**, 39 (1989).
[35] M. Bronzel, *Aktive Beeinflussung nicht-stationärer Schallfelder mit adaptiven Digitalfiltern*, Dissertation, Universität Göttingen (1994).
[36] R. Schirmacher, ‘Application of Fast Adaptive IIR Filter Algorithms for Active Noise Control in a Kitchen Hood’, in *Proc. of Internoise 96* (Liverpool, UK, 1996), pp. 1193–1198.
[37] R. Schirmacher and D. Guicking, ‘Theory and implementation of a broadband active noise control system using a fast RLS algorithm’, Acta Acustica **2**, 291 (1994).
[38] R. Schirmacher, *Schnelle Algorithmen für adaptive IIR-Filter und ihre Anwendung in der aktiven Schallfeldbeeinflussung*, Dissertation, Universität Göttingen (1995).
[39] D. Ronneberger and M. Jüschke, ‘Sound absorption, sound amplification, and flow control in ducts with compliant walls’, in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[40] D. Guicking, ‘Active Control of Sound and Vibration’, in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[41] F. Mechel, ‘Einfluß der Querunterteilung von Absorbern auf die Schallerzeugung in Kanälen’, Acustica **16**, 90 (1965).
[42] D. Ronneberger, G. Höhler, and H. Friedrich, ‘Schallausbreitung in turbulent durchströmten Kanälen mit harten Wänden’, in *Fortschritte der Akustik, DAGA ’76* (1976), pp. 539–542.
[43] C. Ahrens and D. Ronneberger, ‘Luftschalldämpfung in turbulent durchströmten, schallharten Rohren bei verschiedenen Wandrauhigkeiten’, Acustica **25**, 150 (1971).
[44] D. Ronneberger, ‘Experimentelle Untersuchungen zum akustischen Reflexionsfaktor von unstetigen Querschnittsänderungen in einem luftdurchströmten Rohr’, Acustica **24**, 121 (1968).
[45] W. Schilz, ‘Experimentelle Untersuchungen zur akustischen Beeinflussung der Strömungsgrenzschicht in Luft’, Acustica **16**, 208 (1965).
[46] W. Schilz, ‘Untersuchungen über den Einfluß biegeförmiger Wandschwingungen auf die Entwicklung der Strömungsgrenzschicht’, Acustica **15**, 6 (1965).
[47] F. Mechel, W. Schilz, and K. J. Dietz, ‘Akustische Impedanz einer luftdurchströmten Öffnung’, Acustica **15**, 199 (1965).
[48] D. Ronneberger, ‘The Dynamics of Shearing Flow Over a Cavity – A Visual Study Related to the Acoustic Impedance of Small Orifices’, J. Sound Vib. **71**, 565 (1980).
[49] W. Schilz, ‘Richtcharakteristik der Schallabstrahlung einer durchströmten Öffnung’, Acustica **17**, 364 (1966).
[50] D. Ronneberger and U. Ackermann, ‘Experiments on Sound Radiation Due to Nonlinear Interaction of Instability Waves in a Turbulent Air Jet’, J. Sound Vib. **62**, 121 (1979).
[51] D. Ronneberger, ‘Verkehrslärm: Reifenrollgeräusche (Schallschutz im Straßen-

verkehr)', Physik in unserer Zeit **20**, 82 (1989).
[52] F. Evert, D. Ronneberger, and F. R. Grosche, 'Application of linear and nonlinear adaptive filters for the compensation of disturbances in the laminar boundary layer', Z. angew. Math. Mech. **80**, 85 (2000).
[53] A. Ickler, H. Preckel, and D. Ronneberger, 'Observability and Controllability of the Jet-Edge-Flow', in *EUROMECH 415* (Berlin, 2000).
[54] F. Evert, D. Ronneberger, and K. R. Grosche, 'Dynamische Stabilisierung einer laminaren Plattengrenzschicht', Z. angew. Math. Mech. **80**, 603 (1999).
[55] B. Lange and D. Ronneberger, 'Control of Pipe Flow by Use of an Aerodynamic Instability', in *IUTAM Symp. on Mechanics of Passive and Active Flow Control* (Göttingen, 1998), pp. 305–310.
[56] H. Preckel and D. Ronneberger, 'Dynamic Control of the Jet-Edge Flow', in *IUTAM Symp. on Mechanics of Passive and Active Flow Control* (Göttingen, 1998), pp. 349–354.
[57] E. Meyer, W. Kuhl, H. Oberst, E. Skudrzyk, and K. Tamm, 'Sound Absorption and Sound Absorbers in Water', in *Report NavShips 900* (Department of the Navy, Washington, DC, USA, 1950), vol. 1, p. 164.
[58] M. L. Exner, 'Messung der Dämpfung pulsierender Gasblasen in Wasser', Acustica **1**, AB 25 (1951).
[59] E. Meyer and E. Skudrzyk, 'Über die akustischen Eigenschaften von Gasblasenschleiern in Wasser', Acustica **3**, 434 (1953).
[60] R. Mettin, 'From a single bubble to bubble structures in acoustic cavitation', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[61] W. Lauterborn, T. Kurz, R. Geisler, D. Kröninger, and D. Schanz, 'The single bubble – a hot microlaboratory', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[62] E. Meyer, A. Schoch and G. Buchmann, 'Schallschluckanordnung hoher Wirksamkeit', German Patent DE 809 599, filed: July 9, 1938, patented: May 23, 1951; Supplement: DE 878 731, filed: July 1, 1950, patented: Oct. 16, 1952.
[63] E. Meyer and K. Tamm, 'Breitbandabsorber für Flüssigkeitsschall', Acustica **2**, AB91 (1952).
[64] E. Meyer, W. Schilz, and K. Tamm, 'Über den Bau eines reflexionsfreien Wasserschall-Meßbeckens', Acustica **10**, 281 (1960).
[65] D. Guicking and K. Karcher, 'Hydroacoustic Impedance Measurements of Large-Area Samples in the Frequency Range up to 200 kHz', Acustica **54**, 200 (1984).
[66] D. Guicking and T. Wetterling, 'A Reverberation Tank for Hydroacoustic Measurements', in *8th ICA Congress* (London, 1974), p. 482.
[67] D. Guicking and A. Voronovich, 'Theoretical Evaluation of the Edge Effect of an Absorbing Strip on a Pressure-Release Boundary', Acustica **70**, 66 (1990).
[68] J. Richter, H. W. Leuschner, and T. Ahlswede, 'Druckkammerverfahren zur Messung von akustischen Impedanzen in Wasser', Acustica **28**, 90 (1973).
[69] M. Hund, 'Streuung von Wasserschall an zu Biegeschwingungen angeregten Hohlzylindern in einem Flachtank', Acustica **15**, 88 (1965).
[70] P. Wille, 'Experimentelle Untersuchungen zur Schallstreuung an schallweichen Objekten', Acustica **15**, 11 (1965).
[71] H. Peine and D. Guicking, 'Acoustical Resonance Scattering Theory for Strongly Overlapping Resonances', Acta Acustica **3**, 233 (1995).
[72] P. Wille, 'Ein strömungsgünstiges Richtmikrofon für Wasserschall als Analogon des dielektrischen Stielstrahlers', Acustica **17**, 148 (1966).

[73] F. Mechel and P. Wille, ‘Experimentelle Untersuchungen aspektunabhängiger Sonarreflektoren nach dem Prinzip inhomogener Linsen’, Acustica **16**, 305 (1966).
[74] E. Meyer and H. Oberst, ‘Resonanzabsorber für Wasserschall’, Acustica **2**, 149 (1952).
[75] H. Nägerl, ‘Dynamischer Schub- und Elastizitätsmodul amorpher Polymere als Funktion von statischer Vorspannung und Temperatur’, Kolloid-Z. und Z. Polymere **204**, 29 (1965).
[76] D. Guicking, ‘Dynamisch-mechanische Eigenschaften von plastischen Massen und vernetzten Polyurethanen’, Acustica **18**, 93 (1967).
[77] E. Meyer and D. Guicking, *Schwingungslehre* (Vieweg, Braunschweig, 1974).
[78] E. Meyer, K. Brendel, and J. Richter, ‘Absorption von Wasserschall durch erzwungene Wechselströmung von Flüssigkeiten’, Acustica **19**, 8 (1967).
[79] D. Guicking and D. Pallek, ‘Dünnschicht-Resonanzabsorber für Wasserschall’, in *Fortschritte der Akustik – DAGA ’76* (1976), pp. 433–436.
[80] K. Wicker, C. Eberius, and D. Guicking, ‘Electrorheological Fluids as an Electrically Controllable Medium and Possible Applications to Underwater Sound Absorbers’, in *Proc. ACTIVE 97* (1997), pp. 733–744.
[81] D. Guicking, F. Evert et al., ‘Absorberelement zur Absorption von Luftschall’, German Patent DE 196 40 087 C2, filed: Sept. 26, 1996, patented: Jan. 11, 2001.
[82] V. Teubner, ‘Wasserschallabstrahlung von Platten mit und ohne Versteifungen’, Acustica **31**, 203 (1974).
[83] A. W. Leissa, *Vibration of Shells* (NASA Special Publication SP-288, 1973).
[84] D. Guicking and R. Boisch, ‘Vereinfachte Berechnung der Eigenfrequenzen dickwandiger Zylinder in Luft und Wasser’, Acustica **42**, 89 (1979).
[85] D. Guicking and R. Boisch, ‘Zur Grenzfrequenz ebener Platten in dichten Medien’, Acustica **44**, 41 (1980).
[86] E. Meyer and K. Tamm, ‘Eigenschwingung und Dämpfung von Gasblasen in Flüssigkeiten’, Akust. Z. **4**, 145 (1939).
[87] T. Lange, ‘Methoden zur Untersuchung der Schwingungskavitation in Flüssigkeiten mit Ultraschall’, Acustica **2**, 75 (1952).
[88] W. Güth, ‘Kinematographische Aufnahmen von Wasserdampfblasen’, Acustica **4**, 445 (1954).
[89] W. Güth, ‘Nichtlineare Schwingungen von Luftblasen in Wasser’, Acustica **6**, 532 (1956).
[90] W. U. Wagner, ‘Phasenkorrelation von Schalldruck und Sonolumineszenz’, Z. Angew. Physik **10**, 445 (1958).
[91] E. Meyer and K. H. Kuttruff, ‘Zur Phasenbeziehung zwischen Sonolumineszenz und Kavitationsvorgang bei periodischer Anregung’, Z. Angew. Physik **11**, 325 (1959).
[92] K. H. Kuttruff and K. G. Plaß, ‘Sonolumineszenz und Blasenschwingung bei Ultraschallkavitation (30 kHz)’, Acustica **11**, 224 (1961).
[93] W. Lauterborn, ‘Erzeugung und Hochfrequenzkinematographie von Hohlräumen in Wasser mit einem Rubin-Laser’, Research Film **7**, 25 (1970).
[94] W. Lauterborn, ‘Kavitation durch Laserlicht’, Acustica **31**, 51 (1974).
[95] W. Lauterborn, T. Kurz, R. R. Mettin, and C. D. Ohl, ‘Experimental and Theoretical Bubble Dynamics’, in *Advances in Chemical Physics*, edited by I. Prigogine and S. A. Rice (Wiley, New York, 1999), vol. 110, pp. 295–380.
[96] W. Lauterborn, T. Kurz, and I. Akhatov, ‘Nonlinear Acoustics in Fluids’, in *Springer Handbook of Acoustics*, edited by T. D. Rossing (Springer, New York, 2007), chap. 8, pp. 257–297.
[97] W. Lauterborn and U. Parlitz, ‘Methods of Chaos Physics and their Application to Acoustics’, J. Acoust. Soc. Am. **84**, 1975 (1988).

[98] W. Lauterborn and E. Cramer, 'Subharmonic Route to Chaos Observed in Acoustics', Phys. Rev. Lett. **47**, 1445 (1981).
[99] R. Mettin, 'Complex Dynamics of Nonlinear Systems', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[100] D. Krefting, R. Mettin, and W. Lauterborn, 'Two-frequency single-bubble sonoluminescence', J. Acoust. Soc. Am. **112**, 1918 (2002).
[101] W. Lauterborn, T. Kurz, R. Geisler, D. Schanz, and O. Lindau, 'Acoustic cavitation, bubble dynamics and sonoluminescence', Ultrasonics Sonochemistry **14**, 484 (2007).
[102] R. Geisler, W. D. Schmidt-Ott, T. Kurz, and W. Lauterborn, 'Search for neutron emission in laser-induced cavitation', Europhys. Lett. **66**, 435 (2004).
[103] W. Eisenmenger and U. Kaatze, 'Physics of Stone Fragmentation and New Concept of Wide-Focus and Low-Pressure Extracorporal Shock-Wave Lithotripsy', in *Oscillations, Waves, and Interactions* (Universitätsverlag Göttingen, Göttingen, 2007).
[104] A. Vogel, 'UNKNOWN', in *Oscillations, Waves, and Interactions* (Universitätsverlag Göttingen, Göttingen, 2007).
[105] W. Lauterborn, R. Mettin, C. Jung, and R. Sobotta, 'Steuerbare Ultraschallreinigung', in *INFO PHYS TECH* (VDI Technologiezentrum Physikalische Technologien, Düsseldorf, 2002), vol. 36, p. 4 pages.
[106] R. Mettin, J. Appel, D. Krefting, R. Geisler, P. Koch, and W. Lauterborn, 'Bubble structures in acoustic cavitation: observation and modelling of a "jellyfish"-streamer', in *Special Issue of the Revista de Acustica* (Forum Acusticum Sevillia, Sevillia, Spain, 2002), ISBN 84-87985-06-8.
[107] W. Lauterborn, K. Hinsch, and F. Bader, 'Holography of Bubbles in Water as a Method to Study Cavitation Bubble Dynamics', Acustica **26**, 170 (1972).
[108] W. Hentschel and W. Lauterborn, 'Holografische Erzeugung von Kavitationsblasensystemen', in *Fortschritte der Akustik – DAGA '80, München* (VDE-Verlag, Berlin, 1980).
[109] K. J. Ebeling and W. Lauterborn, 'High Speed Holocinematography Using Spatial Multiplexing for Image Separation', Optics Communications **21**, 67 (1977).
[110] K. J. Ebeling and W. Lauterborn, 'Acousto-optic Beam Deflection for Spatial Frequency Multiplexing in High Speed Holocinematography', Applied Optics **17**, 2071 (1978).
[111] K. Hinsch and F. Bader, 'Acoustooptic Modulators as Switchable Beam-Splitters in High-Speed Holography', Optics Communications **12**, 51 (1974).
[112] W. Hentschel and W. Lauterborn, 'New Speed Record in Long Series Holographic Cinematography', Applied Optics **23**, 3263 (1984).
[113] W. Lauterborn and W. Hentschel, 'Cavitation Bubble Dynamics Studied by High Speed Photography and Holography: Part Two', Ultrasonics **24**, 59 (1986).
[114] W. Lauterborn and A. Koch, 'Holographic Observation of Period-Doubled and Chaotic Bubble Oscillations in Acoustic Cavitation', Phys. Rev. A **35**, 1974 (1987).
[115] W. Lauterborn, 'Acoustic Chaos', Physics Today (Jan.) pp. 4–5 (1986).
[116] W. Lauterborn, 'Numerical Investigation of Nonlinear Oscillations of Gas Bubbles in Liquids', J. Acoust. Soc. Am. **59**, 283 (1976).
[117] W. Lauterborn and U. Parlitz, 'Methods of Chaos Physics and their Application to Acoustics', J. Acoust. Soc. Am. **84**, 1975 (1988).
[118] U. Parlitz, V. Englisch, C. Scheffczyk, and W. Lauterborn, 'Bifurcation structure of bubble oscillators', J. Acoust. Soc. Am. **88**, 1061 (1990).
[119] 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. **107A**, 351 (1985).

[120] C. Scheffczyk, U. Parlitz, T. Kurz, W. Knop, and W. Lauterborn, ‘Comparison of bifurcation structures of driven dissipative nonlinear oscillators’, Phys. Rev. A **43**, 6495 (1991).
[121] R. Mettin, U. Parlitz, and W. Lauterborn, ‘Bifurcation Structure of the Driven Van der Pol Oscillator’, Int. J. Bif. Chaos **3**, 1529 (1993).
[122] K. Geist and W. Lauterborn, ‘The Nonlinear Dynamics of the Damped and Driven Toda Chain: II. Fourier and Lyapunov analysis of tori’, Physica D **41**, 1 (1990).
[123] U. Dressler and W. Lauterborn, ‘Ruelle’s rotation frequency for a symplectic chain of dissipative oscillators’, Phys. Rev. A **41**, 6702 (1990).
[124] W. Lauterborn and R. Steinhoff, ‘Bifurcation Structure of a Laser with Pump Modulation’, J. Opt. Soc. Am. B **5**, 1097 (1988).
[125] V. Ahlers, U. Parlitz, and W. Lauterborn, ‘Hyperchaotic Dynamics and Synchronization of External-Cavity Semiconductor Lasers’, Phys. Rev. E **58**, 7208 (1998).
[126] A. Ahlborn and U. Parlitz, ‘Laser stabilization with multiple delay feedback control’, Opt. Lett. **31**, 465 (2006).
[127] T. Klinker and W. Lauterborn, ‘Scattering of Lattice Solitons from a Mass Interface – a Synergetic Approach’, Physica D **8**, 249 (1983).
[128] A. Kumar, T. Kurz, and W. Lauterborn, ‘Two-state bright solitons in doped fibers with saturating nonlinearity’, Phys. Rev. E **53**, 1166 (1996).
[129] E. Meyer and H. Severin, ‘Über einige Demonstrationsversuche mit elektromagnetischen Zentimeter-Wellen’, Z. Physik **126**, 711 (1949).
[130] H. Severin and W. von Baeckmann, ‘Beugung elektromagnetischer Zentimeterwellen an metallischen und dielektrischen Scheiben’, Z. Angew. Physik **3**, 22 (1951).
[131] H. Severin and K. Körper, ‘Beugung elektromagnetischer Wellen an rechteckigen Öffnungen in ebenen Metallschirmen’, Z. Angew. Physik **13**, 41 (1960).
[132] E. G. Neumann, ‘Über das elektromagnetische Feld der schwach geführten Dipolwelle’, Z. Angew. Physik **16**, 452 (1964).
[133] E. G. Neumann, ‘Experimentelle Untersuchung der Oberflächenwelle an einer Yagi-Struktur’, Z. Angew. Physik **19**, 121 (1965).
[134] H. J. Schmitt, ‘Breitbandiger Resonanzabsorber für elektromagnetische Zentimeterwellen’, Z. Angew. Physik **8**, 373 (1956).
[135] G. Kurtze and E. G. Neumann, ‘Ein Dipolabsorber für elektromagnetische Zentimeterwellen mit verminderter Reflexion bei schräger Inzidenz’, Z. Angew. Physik **12**, 385 (1960).
[136] H. W. Helberg and C. Wünsche, ‘Ein Mehrschichtabsorber für elektromagnetische Zentimeterwellen’, Z. Angew. Physik **16**, 157 (1963).
[137] W. Burgtorf and H. P. Seraphim, ‘Eine Apparatur zur elektroakustischen Herstellung einfacher Schallfelder in einem reflexionsfreien Raum’, Acustica **11**, 92 (1961).
[138] C. Wünsche, ‘Nichtlinear dielektrisches Verhalten von Ferroelektrika im Mikrowellengebiet bei 9,4 GHz’, Z. Angew. Physik **19**, 501 (1965).
[139] V. Kose, ‘Feldabhängigkeit der komplexen Dielektrizitätskonstante von Bariumtitanat bei 58,2 GHz’, Z. Angew. Physik **23**, 425 (1967).
[140] U. Kaatze, ‘Messungen der komplexen Dielektrizitätszahl von Bariumtitanat-Einkristallen bei Mikrowellenfrequenzen’, Phys. Stat. Sol. **50**, 537 (1972).
[141] K. D. Toepfer and H. W. Helberg, ‘Dielectric Dispersion of Thiourea Single Crystals at Microwave Frequencies between 8.25 and 75.0 GHz’, Phys. Stat. Sol. **35**, 131 (1976).
[142] H. W. Helberg and V. Kose, ‘Die breitbandige Absorption elektromagnetischer Wellen durch dünne Ferritschichten’, Z. Angew. Physik **19**, 509 (1965).
[143] K. Giese, ‘Über ferromagnetische Resonanz in Kugeln aus polykristallinen Ferriten’, Z. Angew. Physik **23**, 440 (1967).

[144] R. Pottel and A. Wülfing, 'Messungen der komplexen Dielektrizitätskonstante von wasserhaltigem Glycerin und Glycerin-Gelatine-Gelen bei Frequenzen zwischen 100 MHz und 15 GHz', Z. Angew. Physik **15**, 501 (1963).
[145] K. G. Plaß, 'Relaxationen in organischen Flüssigkeiten bei 1 GHz', Acustica **19**, 236 (1967).
[146] R. Pottel, J. Haller, and U. Kaatze, 'Multistep Association of Cations and Anions. The Eigen-Tamm Mechanism Some Decades Later', in *Oscillations, Waves, and Interactions* (Universitätsverlag Göttingen, Göttingen, 2007).
[147] U. Kaatze and R. Behrends, 'Liquids: Formation of Complexes and Complex Dynamics', in *Oscillations, Waves, and Interactions* (Universitätsverlag Göttingen, Göttingen, 2007).
[148] R. Behrends and U. Kaatze, 'A High Frequency Shear Wave Impedance Spectrometer for Viscosity Liquids', Meas. Sci. Technol. **12**, 519 (2001).
[149] C. Trachimow, L. De Maeyer, and U. Kaatze, 'Extremely Slow Reaggregation Processes in Micelle Solutions. A Dynamic Light Scattering Study', J. Phys. Chem. B **102**, 4483 (1998).
[150] R. Pottel, 'Die komplexe Dielektrizitätskonstante wäßriger Lösungen einiger 2-2-wertiger Elektrolyte im Frequenzbereich 0,1 bis 38 GHz', Ber. Bunsenges. f. Physikal. Chemie **69**, 363 (1965).
[151] K. Tamm and G. Kurtze, 'Absorption of Sound in Aqueous Solutions of Electrolytes', Nature **168**, 346 (1951).
[152] M. Eigen, G. Kurtze, and K. Tamm, 'Zum Reaktionsmechanismus der Ultraschallabsorption in wäßrigen Elektrolytlösungen', Ber. Bunsenges. f. Physikal. Chemie **57**, 103 (1953).
[153] U. Stumper, 'Dielectric Absorption of Liquid Normal Alkanes in the Microwave and Far Infrared Regions', Advances in Molecular Relaxation Processes **7**, 189 (1975).
[154] O. Göttmann, 'Dielektrische Relaxation in flüssigen binären Lösungen nichtdipolarer Elektronen-Donator- und Elektronen-Akzeptor-Moleküle', Ber. Bunsenges. f. Physikal. Chemie **80**, 280 (1976).
[155] R. Behrends and U. Kaatze, 'Structural Isomerization and Molecular Motions of Liquid n-Alkanes. Ultrasonic and High-Frequency Shear Viscosity Relaxation', J. Phys. Chem. A **104**, 3269 (2000).
[156] W. Eisenmenger, H. Kinder, and K. Laßmann, 'Messung der Hyperschalldämpfung in Quarz', Acustica **16**, 1 (1965).
[157] W. Eisenmenger, 'Erzeugung und Nachweis von höchstfrequentem Schall durch Quantenprozesse in Supraleitern', Nachrichten der Akademie der Wissenschaften in Göttingen **24**, 115 (1966).
[158] W. Eisenmenger and A. H. Dayem, 'Quantum Generation and Detection of Incoherent Phonons in Superconductors', Phys. Rev. Lett. **18**, 125 (1967).
[159] M. Dressel, 'Charge-Ordering Phenomena in One-Dimensional Solids', in *Oscillations, Waves, and Interactions* (Universitätsverlag Göttingen, Göttingen, 2007).
[160] H. W. Helberg and B. Wartenberg, 'Die elektrische Leitfähigkeit von pyrolysiertem Polyacrylnitril im Temperaturbereich 1,7 bis 700 K', Phys. Stat. Sol. **3**, 401 (1970).
[161] M. Przybylski and H. W. Helberg, 'Frequency-Dependent Transport in the Pure and Irradiation-Disordered Organic Semiconductor N-Methyl- (4-Methyl) Pyridinium (7,7,8,8-Tetracyano-p-Quinodimethanide)', Phys. Rev. B **31**, 8034 (1985).
[162] H. W. Helberg, 'Electronic Excitations in Alpha-(BEDT-TTF)$_3$(NO_3)$_2$', Physica B **143**, 488 (1986).
[163] H. W. Helberg and M. Dressel, 'Investigations of Organic Conductors by the Schegolev Method', J. de Physique I France **6**, 1683 (1996).

[164] P. Polanowski, J. Ulanski, R. Wojciechowski, A. Tracz, J. K. Jeszka, S. Matejcek, E. Dormann, B. Pongs, and H. W. Helberg, 'Thin Layers of ET_2I_3 Obtained by IN SITU Crystallization – the Role of Polymer Matrix', Synthetic Metals **102**, 1789 (1999).

[165] S. Halstenberg, W. Schrader, P. Das, J. K. Bhattacharjee, and U. Kaatze, 'Critical Fluctuations in the Domain Structure of Lipid Membranes', J. Chem. Phys. **118**, 5683 (2003), also: Virtual J. Biol. Phys. Res. 5 (March 15, 2003).

[166] W. Schrader, S. Halstenberg, R. Behrends, and U. Kaatze, 'Critical Slowing in Lipid Bilayers', J. Phys. Chem. B **107**, 14457 (2003).

[167] V. Oliynyk, U. Kaatze, and T. Heimburg, 'Defect formation of lytic peptides in lipid membranes and their influence on the thermodynamic properties of the pore envirement', Biochimica et Biophysica Acta **1768**, 236 (2007).

[168] R. Polacek, R. Behrends, and U. Kaatze, 'Chair-Chair Conformational Flexibility of Monosaccharides Linked to the Anomer Equilibrium', J. Phys. Chem. B **105**, 2894 (2001).

[169] R. Behrends and U. Kaatze, 'Molecular Dynamics and Conformational Kinetics of Mono- and Disaccharides in Aqueous Solution', ChemPhysChem **6**, 1133 (2005).

[170] U. Kaatze, R. Henze, and H. Eibl, 'Motion of the Lengthened Zwitterionic Heat Groups of C_{16}-Lecithin Analogues in Aqueous Solutions as Studied by Dielectric Relaxation Measurements', Biophys. Chem. **10**, 351 (1979).

[171] R. Henze, E. Neher, T. L. Trapane, and D. W. Urry, 'Dielectric Relaxation Studies of Ionic Processes in Lysolecithin-Packaged Gramicidin Channels', J. Membrane Biol. **64**, 233 (1982).

[172] R. Behrends, M. K. Cowman, F. Eggers, E. M. Eyring, U. Kaatze, J. Majewski, S. Petrucci, K. H. Richmann, and M. Riech, 'Ultrasonic Relaxation and Fast Chemical Kinetics of Some Carbohydrate Aqueous Solutions', J. Am. Chem. Soc. **119**, 2182 (1997).

[173] S. Halstenberg, T. Heimburg, T. Hianik, U. Kaatze, and R. Krivanek, 'Cholesterol Induced Variations in the Volume and Enthalpy Fluctuations of Lipid Bilayers', Biophys. J. **75**, 264 (1998).

[174] J. Stenger, M. Cowman, F. Eggers, E. M. Eyring, U. Kaatze, and S. Petrucci, 'Molecular Dynamics and Kinetics of Monosaccharides in Solution. A Broadband Ultrasonic Relaxation Study', J. Phys. Chem. B **104**, 4782 (2000).

[175] D. Schild and D. Guicking, 'A Novel Approach to the van der Pol Oscillator: Natural Frequency and Entrainment', J. Interdiscipl. Cycle Res. **11**, 285 (1980).

[176] R. Pottel and A. Protte, 'Surface Transmission Probe for Noninvasive Measurements of Dielectric Properties of Organ Tissues at Frequencies between 1 MHz and 300 MHz', Biomed. Technik **35**, 158 (1990).

[177] E. Gersing, B. Hofmann, G. Kehrer, and R. Pottel, 'The Modelling of Cellular Media in Electrical Impedance Tomography', Innov. Tech. Biol. Med. **16**, 671 (1995).

[178] A. Vogel, W. Hentschel, J. Holzfuss, and W. Lauterborn, 'Kavitationsblasendynamik und Stoßwellenabstrahlung bei der Augenchirurgie mit gepulsten Neodym:YAG-Lasern', Klin. Mbl. Augenheilk. **189**, 308 (1986).

[179] A. Vogel, W. Hentschel, J. Holzfuss, and W. Lauterborn, 'Cavitation Bubble Dynamics and Acoustic Transient Generation in Ocular Surgery with Pulsed Neodymium:YAG Lasers', Ophthalmology **93**, 1259 (1986).

[180] B. Wolfrum, R. Mettin, T. Kurz, and W. Lauterborn, 'Observation of pressure-wave-excited contrast agent bubbles in the vicinity of cells', Appl. Phys. Lett. **81**, 5060 (2002).

[181] B. Wolfrum, T. Kurz, R. Mettin, and W. Lauterborn, 'Shock wave induced interaction

of microbubbles and boundaries', Phys. Fluids **15**, 2916 (2003).
[182] P. Zwirner, D. Michaelis, and E. Kruse, 'Akustische Stimmanalysen zur Dokumentation der Stimmrehabilitation nach laserchirurgischer Larynxkarzinomresektion', HNO **44**, 514 (1996).
[183] D. Michaelis, M. Fröhlich, and H. W. Strube, 'Selection and combination of acoustic features for the description of pathologic voices', J. Acoust. Soc. Am. **103**, 1628 (1998).
[184] C. Rotsch, K. Jacobson, J. Condeelis, and M. Radmacher, 'EGF-stimulated lamellipod extension in adenocarcinoma cells', Ultramicroscopy **86**, 97 (2001).
[185] R. Matzke, K. Jacobson, and M. Radmacher, 'Direct, high-resolution measurement of furrow stiffening during division of adherent cells', Nature Cell Biol. **3**, 607 (2001).
[186] B. H. Kwok, L. C. Kapitein, J. H. Kim, E. J. G. Peterman, C. F. Schmidt, and T. M. Kapoor, 'Allosteric Inhibition of Kinesin-5 Modulates its Processive Directional Motility', Nature Chem. Biol. **2**, 480 (2006).
[187] M. Korneev, S. Lakämper, and C. F. Schmidt, 'Load-Dependent Release Limits the Progressive Stepping of the Tetrameric Eg5 Motor', Eur. Biophys. J. (2007).
[188] D. Mizuno, C. Tardin, C. F. Schmidt, and F. C. MacKintosh, 'Nonequilibrium Mechanics of Active Cytoskeletal Networks', Science **315**, 370 (2007).
[189] C. Schmidt and S. Lakämper, 'UNKNOWN', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).

Photograph "Free-space" room: LIFE, International Edition, May 31, 1954.
Photographs and image processing: Gisa Kirschmann-Schröder

Oscillations, Waves and Interactions, pp. 25–36
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-02-4

Speech research with physical methods

Hans Werner Strube

Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. An overview of some recent work in speech research at the Dritte Physikalische Institut is given, especially of investigations from a cooperation between physics and phoniatrics that concern the analysis of pathologic voices by acoustic and optical means. The main novel points are the extension of our own previously published acoustic methods to running speech and new high-speed video methods.

1 Overview

Recent work at the Dritte Physikalische Institut may be divided in two thematic fields:

- Work related to speech recognition.
- Acoustic analysis of pathologic voices, extended to running speech.

Here only the second thematic field, which was carried out as a cooperative project of Prof. Eberhard Kruse (Department of Phoniatrics and Paedaudiology, University of Göttingen) and our group, will be described in more detail.

2 Work related to speech recognition

This research concerned, on one hand, methods appropriate for preprocessing in speech recognition, such as novel approaches to noise reduction, employing filtering in the modulation-frequency domain [1,2], and to speaker normalization, starting from acoustic estimation of speaker-specific measures of the vocal tract [3]. On the other hand, there were recent investigations concerning speech recognition itself: first, Hidden Markov Model (HMM) based recognition for "endless" signals with continuous forming of hypothesis graphs [4,5] and noise-robust speech/nonspeech distinction based on modulation filtering [6] (partially carried out at DaimlerChrysler); second, exploitation of prosodic features (measures of pitch and loudness) to improve semantic recognition in the context of a natural speech dialog platform [7,8] (partially done at Bosch GmbH).

3 Acoustic analysis of pathologic voices, extended to running speech[1]

3.1 Introduction

For diagnostics and treatment of voice disorders the evaluation of voice quality is essential. Apart from the auditive judgment by the phoniatrist, it is desirable to find objective criteria for the rating of voice disorders, especially, to determine physical quantities appropriate for diagnostic description and classification of speech pathologies. For this purpose, acoustic analysis methods have been developed. Starting from known quantities, such as jitter, shimmer, and measures of additive noise, novel quantities were investigated that allow a better separation of noise from glottal periodicity disturbances. According to clinical usage, first stationary vowels were considered. But an analysis of running speech is also desirable, in order to assess the voice under more realistic conditions. So an important goal of the project was the extension of the methods to running speech.

Further, in phoniatric diagnostics stroboscopic video recordings of the vibrating glottis are usual. These were supplemented by high-speed recordings, and automatic methods of image segmentation were developed (e. g., determination of the glottal opening area). Besides all this, an extensive data bank with acoustic and optical recordings as well as diagnostic findings was built up.

3.2 Equipment and data bank

For recording and processing, standard PCs were used with good sound cards (e. g., Soundblaster PCI 128) under Linux in an ethernet LAN. The voice recordings were done in an anechoic and insulated recording room, not containing any noisy devices. The microphone, a Beyerdynamic HEM 191.15 with spherical characteristic, was head-fixed about 10 cm in front of the chin. A special graphic interface for recording, cutting and marking of the voice recordings was programmed.

Presently there are about 77000 voice recordings as WAV files with 48 kHz sampling frequency: vowels [ɛ: a: e: i: o: u:] with pitches normal, low, high; running speech (phonetically balanced standard texts “Nordwind und Sonne” [north wind and sun] and “Buttergeschichte” [butter story]); spontaneous speech. For archiving and automatic administration of the voice and video recordings as well as the medical diagnoses (about 70 different ones), a large MySQL data base was built up. It runs with a PHP web frontend on a Linux PC, is connected to the patient information system SAP of the university hospital and has an interface to the video stroboscopy workplaces. For each patient-related voice analysis, a PDF file with color print can be generated.

3.3 The Göttinger Hoarseness Diagram

Especially fruitful was the voice characterization by the “Göttinger Hoarseness Diagram” (GHD) [11–14], Fig. 1. It is based on traditionally important quantities in phoniatric diagnostics: on one hand, irregularity measures of glottal oscillation, such

[1] extended and updated from earlier German papers [9,10]

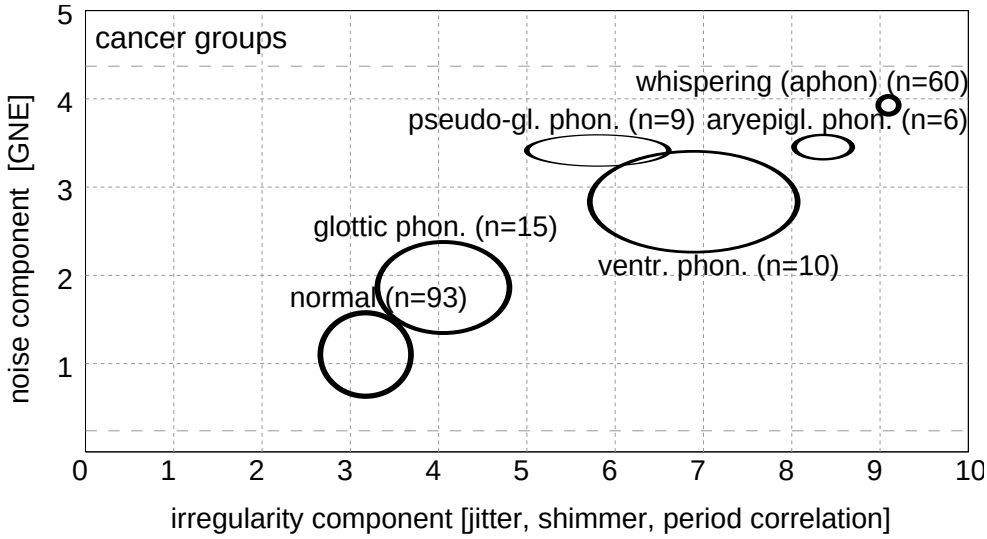


Figure 1. Göttinger Hoarseness Diagram; distribution of some cancer groups; cf. Ref. [13].

as jitter (period-length fluctuation) and shimmer (amplitude or energy fluctuation); on the other hand, measures of the noise component relative to glottal excitation. These are coarse correlates of subjective roughness and breathiness, respectively. There are many different definitions of such measures; e. g., for irregularity [15]:

$$\left|\frac{x_n - x_{n-1}}{x_n}\right| \quad \text{or} \quad \left|\frac{x_n - (1/M)\sum_{m=-(M-1)/2}^{(M-1)/2} x_{n+m}}{(1/M)\sum_{m=-(M-1)/2}^{(M-1)/2} x_{n+m}}\right|, \; M = 3, 5, \ldots ,$$

each averaged over n, where x_n is the length or amplitude or energy of the nth period. The period energy is more robustly measurable than the amplitude. For the period length, a method proved to be especially reliable which was based on the correlation coefficient of subsequent signal intervals $[t,\, t+T)$ and $[t+T,\, t+2T)$ and maximizing with respect to T [16,17] (with interpolation between the signal samples). The average correlation coefficient of subsequent periods served as another irregularity measure (Mean Waveform matching Coefficient, MWC).

Traditional measures of the noise component (e. g., NNE [18], CHNR [19]) are unfortunately dependent on the irregularity measures and the choice of the analysis window. With strongly irregular voices, they are often not applicable, since they usually require a harmonic spectral structure. Therefore, we started from the following assumption [20]: with glottal excitation – regardless how irregular – the excitation in different bands should be nearly synchronous, with noise excitation, however, asynchronous. Hereon the following construction is based. After downsampling to 10 kHz and linear-predictive inverse filtering for removing the formant structure, the signal is decomposed into partial bands using Hann-window shaped filters. For each band, the Hilbert envelope is formed and its mean removed. For all pairs of bands

that overlap at most half, normalized cross-correlation coefficients are then formed with shifts within ±0.3 ms. The largest of all these coefficients yields a meassure, "Glottal-to-Noise Excitation ratio" (GNE) [20].

To obtain a graphical representation appropriate for clinical routine, a two-dimensional plot was desired. For this purpose we considered 6 jitter and 6 shimmer measures, the MWC, and 3 variants each of NNE, CHNR and GNE. The dependencies between these were investigated using rank correlation and mutual information. It turned out that the GNE was most independent of the irregularity measures. By means of principal-component analysis it was shown that, for pathological voices, two dimensions explain 95% of variance (for normal voices, four dimensions are required). In this way, a diagram resulted with an abscissa that was an average of a jitter measure, a shimmer measure and the MWC, and a (linearly transformed) GNE as ordinate, the "Göttinger Hoarseness Diagram" (GHD).

Normal voices are located at the lower left, aphonic voices at the upper right in the diagram. Different groups (e. g., persons, vowels, medical diagnoses) can be represented by ellipses, where the principal axes indicate the standard deviations with respect to abscissa and ordinate. Fig. 1 shows an example for various cancer groups.

As in the regions of normal and of aphonic voices two dimensions are not really sufficient, it was also tried to obtain a finer resolution using Kohonen feature maps [21] (example see Fig. 2) and discriminant analysis [22]. Also, the relevance of several breathiness measures for classifying voice pathologies was investigated [23], resulting in a dimensionality of 7 for benign disorders and 3 for cancer groups.

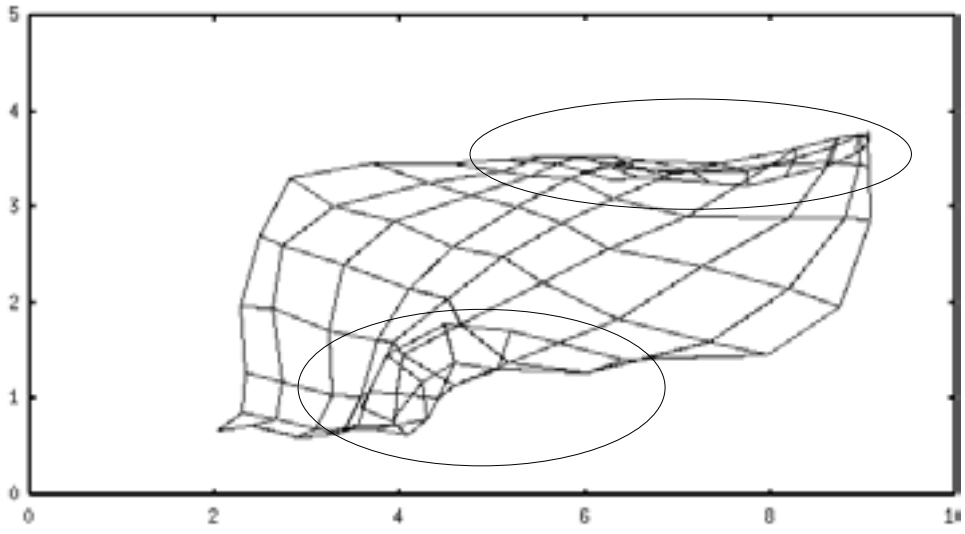


Figure 2. Kohonen feature map, projected into the GHD plane. Note the folds (marked by ellipses) near the regions of normal and aphonic voices, indicating importance of higher dimensions.

3.4 Transfer to running speech

The voice analysis was originally based on stationary vowels. In clinical diagnostics, however, a voice analysis from continuous speech is required in order to objectively assess the vocal disease under normal stress and to be able to treat it optimally. The stationary phonation corresponds rather to a singing voice, contrary to the more natural running speech. Thus for a comprehensive description of voice quality the analysis of running speech is an essential extension of the analysis of stationary phonation. The methods of vowel analysis should be partially transferable to voiced intervals in running speech. For this purpose a method was developed to recognize such intervals automatically. The main difficulty was that the linguistically voiced sounds are not necessarily realized as voiced for strong voice disorders.

3.4.1 Determination of voiced and unvoiced intervals

A voiced/unvoiced classification by (e. g.) zero-crossing and correlation techniques directly on the speech signal would, for strongly disturbed voices, recognize too few voiced intervals. Thus a consideration of the spectral envelope (formant structure) is preferable, which little depends on the actual glottal excitation. The method uses a 3-layer perceptron with sigmoid activation function (values 0 to 1) as classifier. The template vectors for its input were formed as follows (numbers refer to Fig. 3):
The speech signals, digitized with 48 kHz, were downsampled to 12 kHz and decomposed into overlapping Hann-windowed 40 ms intervals with 10 ms frame shift (3,4). Pauses are eliminated based on an empirical energy threshold. An LPC analysis of 12^{th} order (autocorrelation method; preemphasis 0.9735) yields a model spectrum (5), which is converted to 19 critical bands (Bark scale) by summation in overlapping trapezoidal windows (6). It is compressed with exponent 0.23 and normalized by its maximum over time and critical bands (7,8). The LPC order and method were optimized to yield minimal misclassification.

The optimal values of the perceptron parameters (number of hidden cells, learning rate, classification threshold, number of iterations, training material) were determined in extensive experiments (6750 different cases, about 12000 spectra). Twelve hidden cells worked best. As training method of the perceptron (9), an accelerated backpropagation [24] was employed with learning rate 0.01 and momentum term 0.8. The classification threshold at the output is 0.45, the desired net outputs for training are 0.1 (unvoiced) and 0.9 (voiced). The weights are initialized with random numbers in the range 0 to 1. Three perceptrons with different initial weights were used in parallel, averaging their recognition scores.

Since our own speech data were unlabeled, they could not serve for training. Instead, the training set consisted first of 32 phonetically segmented texts (16 times "Nordwind und Sonne", 16 times "Berlingeschichte") from 16 different normal speakers in the German Phondat database, a total of 154550 labeled Bark spectra, excluding pauses. The training used up to 5000 iterations. For testing, different subsets of 30 of the 32 texts were used in training and the remaining 2 in the test. The error score amounted to 4.8%. With the above threshold (0.45), only 25% of these are falsely classified as voiced; false unvoiced classification is less detrimental. As mis-

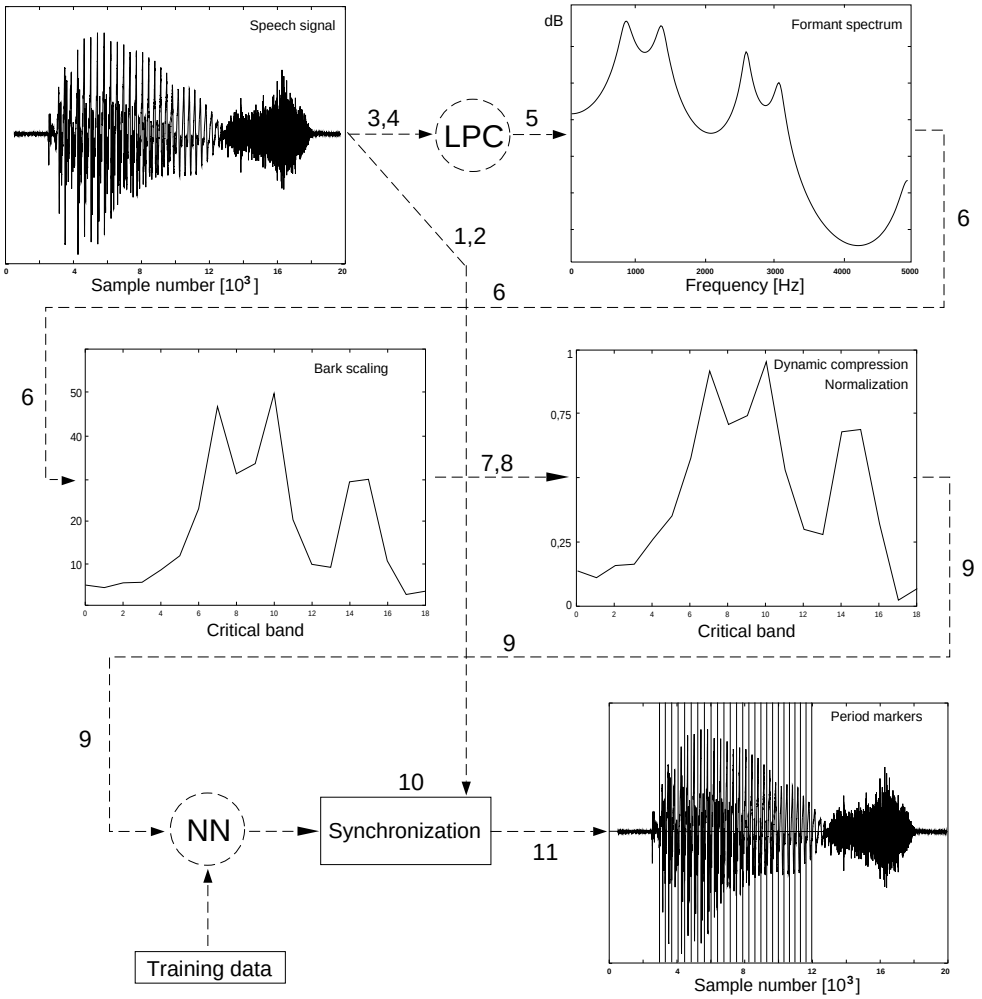


Figure 3. Voiced/unvoiced classification, see text; cf. Ref. [25]

classification increased near phoneme boundaries and voiced/unvoiced boundaries, the first and last frame of a segment of equal phonation class were discarded and plosives were always treated as unvoiced.

After this, a post-training with 12500 spectra from pathologic speech had to be performed. The learning rate was chosen smaller to avoid too much unlearning of the previous training with normal voices. For the voiced intervals, only eight stationary vowels were used here, but for the unvoiced ones, manually selected consonants from running speech ("Nordwind und Sonne").

3.4.2 Analysis of running speech

The analysis was only carried out on connected intervals classified as voiced of minimum length 70 ms. A weighting by length is implemented, since longer intervals are more expressive. In these intervals period markers are set (Fig. 3 (11)) and acoustic quantities are determined, for instance:

- period lengths by the waveform matching algorithm;
- jitter (3 definitions), shimmer (3 definitions), MWC, GNE.

From these again a GHD can be constructed. The position of voices in the GHD is different in running speech from that for stationary vowels, so that a new calibration is required to obtain comparable representations for both cases. The definition of the axes, which is based on a principal-component analysis in a high-dimensional space, has to be carried out anew. Here, the choice of the underlying quantities was the same for consistency reasons, but their weighting was different. The new GHD is called "GHDT", "T" meaning "text". The coordinates in the GHDT are averaged over the analyzed intervals of the text utterance, weighted by their lengths. The variances of the measurement points in the GHD are, because of sound dependence, of course larger than for stationary vowels, but the mean values retain their expressiveness. The consistency of the GHDT was checked with various normal and pathological voices and different utterances.

Besides the GHD, the automatic voiced/unvoiced classification can also be applied to other diagnostically useful quantities in order to extend their usage to running speech. This concerns, for instance, the Pitch Amplitude (PA; 1^{st} maximum of the autocorrelation function of the prediction error signal) and the Spectral Flatness Ratio (SFR; logarithm of the ratio of geometric and arithmetic means of the spectral energy density of the prediction error signal).

Based on the acoustic quantities, group analyses of various phonation mechanisms and cancer groups (significant group separation) can be conducted. For preliminary and recent presentations of methods and results see Refs. [25–28].

So far, no phonemes were to be recognized but only their linguistic (not actual) voicedness. Meanwhile, the perceptron method has been extended to recognition of the six stationary vowels, using 6 output cells. Training was done with 8192 vowels of at least 2 s duration from all kinds of voice quality. This can help to further automatize the determination of voice quality.

3.5 Analysis of glottal oscillation

The voice pathologies are related to the functioning of the vocal folds, which form a self-oscillating nonlinear mechanic and aerodynamic system driven by the glottal air flow. In order to relate the acoustic voice characteristics to properties of the glottal oscillation, these must be (if possible, automatically) recorded and characterized by few quantities. Here, acoustic as well as optical methods are employed. These methods have not yet been extended to running speech, but the only essential difficulty to do so appears to be the large amount of data occurring then.

3.5.1 Glottis pulse parameters by inverse filtering

For describing the glottal oscillation, it is hardly feasible to estimate the many mechanical parameters of the vocal-fold system, but characteristic quantities can be used as occurring in parametric models of the pulse shape of the glottal volume velocity, for instance, the well-known LF-model [29]. Such a model can be fitted to a measured pulse shape. For the measurement to be feasible during normal speaking, only inverse-filtering techniques are applicable. That is, the filtering by the vocal tract is computationally undone in order to obtain the glottal volume velocity. This leads to the circular problem that the vocal-tract transfer function cannot be estimated exactly enough without knowledge of the input signal. Thus, a simultaneous estimation of the transfer function and the glottal pulse (as LF-model) was carried out iteratively with a multidimensional optimization method, see Fig. 4.

The transfer function was computed in a pitch-related way with a modified DAP algorithm [32] and the Itakura-Saito error occurring therein was also used as optimality criterion for the total iteration. The method was verified with synthetic speech, where beside the LF parameters also derived quantities such as Open-Quotient, Closed-Quotient, Speed-Quotient, Parabolic Spectral Parameter were compared with the given data. For natural speech, the energies of 200-ms intervals and the electroglottographically measured open-quotient were employed for comparison with the model. A detailed description can be found in Refs. [30,31].

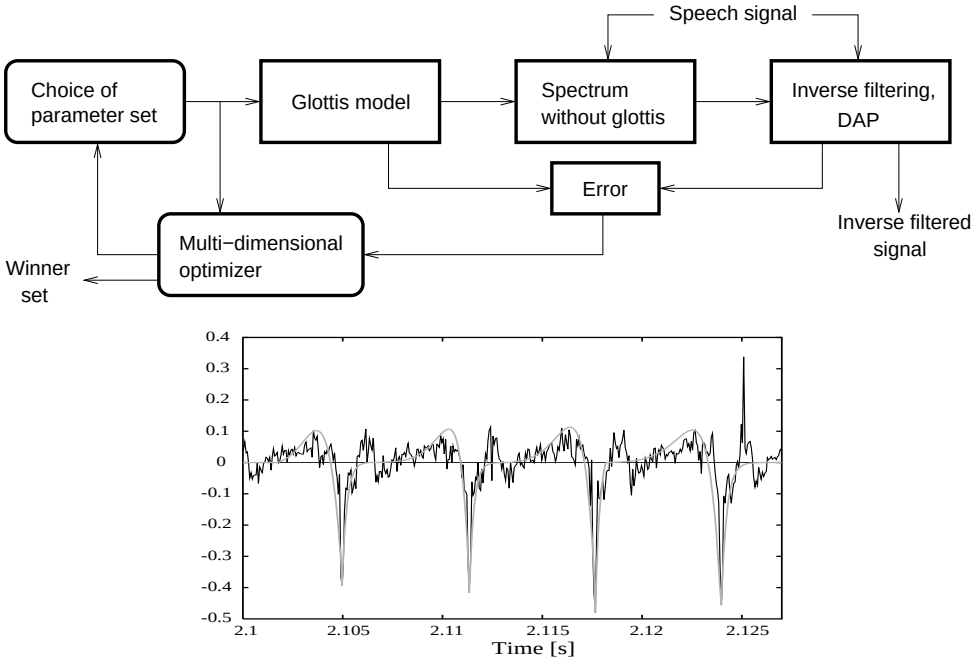


Figure 4. Top: Scheme of estimating glottal pulse and transfer function, cf. [30,31]. Bottom: Example of estimated glottal flow derivative and fitted LF-model (gray, smooth).

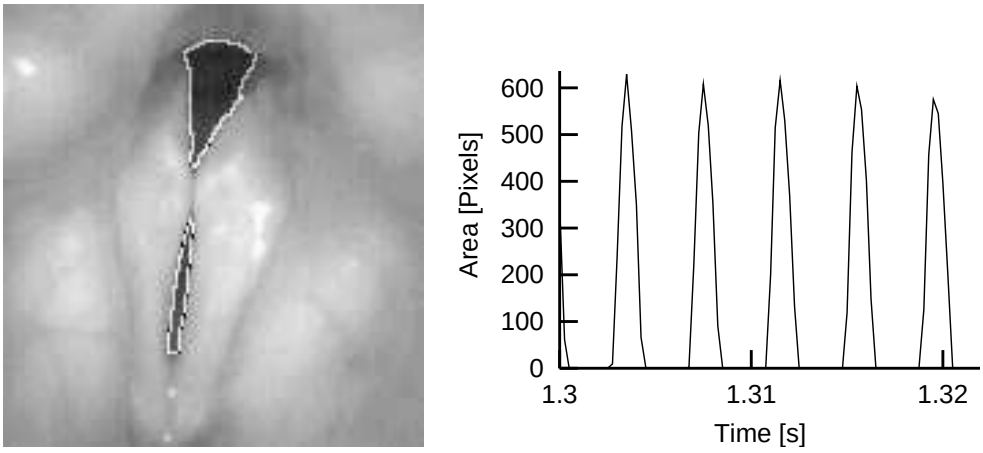


Figure 5. Left: disconnected glottal area and its recognized boundary (white). Right: Example of a measured opening-area time course. After Ref. [33].

3.5.2 Video evaluations

Although the video methods are not presently applicable to running speech, they must be mentioned here because of their importance. The acoustic measurable quantities are to be related to properties of the glottal oscillation.For automatic evaluation of the stroboscopic color recordings usual in phoniatric diagnostics an automatic image segmentation method was developed that recognizes objects such as vocal folds, glottal opening etc. using a perceptron.

The details of the oscillations were investigated by means of high-speed recordings (Wolf HS ENDOCAM 5560, 8-bit grayscale, 256 × 256 pixels, 4000 frames/s, max. 8192 frames). The glottal opening was automatically recognized by a three-dimensional (area × time) algorithm [33] (older version see Ref. [34]). First, "active regions" are determined that show fast temporal brightness changes, and the largest one is circumscribed by a rectangle. In this three-dimensional rectangle×time region, an inner and an outer region are constructed in the following way. Starting from seed points appropriately chosen for each time instant, a competitive region growth (modified after Ref. [35]) is performed into neighborhoods chosen to be as homogeneous as possible, until both regions border on each other. The inner region is slightly enlarged by erosion [35]. Even disconnected glottal areas are thus correcty recognized, Fig. 5.

For investigating the relation of glottal quantities to acoustic quantities (e. g., GNE), from 15 recordings 263 data sets of 100 ms each (sufficiently voiced, not overlapping) were constructed. Each contained the four values GNE, mean maximum glottal area, mean minimum glottal area and closed-quotient (CQ, relative time of closed glottis). There was a clear correlation between CQ and GNE (0.536, nonzero with significance level $3.3 \cdot 10^{-18}$), likewise between the minimum opening area and the GNE (–0.464, nonzero with significance level $5.1 \cdot 10^{-14}$). Spearman rank correlation coefficients were also tried and were close to these values. This corresponds to

the expectations: the longer and tighter the glottal closure is, the smaller the noise component should be.

As for the extension to running speech, the high-speed video method is limited by the amount of frames that can be recorded continuously, about 2 s of speech.

Acknowledgements. The author thanks Prof. Eberhard Kruse (Dept. of Phoniatrics and Paedaudiology) for his engaged project management and all of our coworkers, especially Matthias Fröhlich, Dirk Michaelis, Jan Lessing, Sven Anderson, who have mainly carried out the work described here. We are indebted to the Deutsche Forschungsgemeinschaft (DFG) for funding this work under grants Kr 1469/2 and 1469/5.

References

[1] H. W. Strube and H. Wilmers, 'Noise reduction for speech signals by operations on the modulation frequency spectrum', in *Joint Meeting "Berlin 1999" ASA/EAA/DEGA* (DEGA, Oldenburg, 1999), full paper on CD-ROM; Abstracts in: Acustica / acta acustica 85, S52 (1999) and J. Acoust. Soc. Am. **105**, 978 (1999).

[2] O. Schreiner and H. W. Strube, 'Modulationsfilterung von Sprache mit Fourier-Spektrogramm und Wavelet-Transformation', in *Fortschritte der Akustik – DAGA 2001*, edited by O. von Estorff (DEGA, Oldenburg, 2001), pp. 100–101.

[3] H. Freienstein, K. Müller, and H. W. Strube, 'Bestimmung von sprecherspezifischen Vokaltraktparametern', in *Elektronische Sprachsignalverarbeitung. Tagungsband der zehnten Konferenz*, edited by D. Mehnert (w.e.b. Universitätsverlag, Dresden, 1999), vol. 16 of *Studientexte zur Sprachkommunikation*, pp. 208–215.

[4] C. Rico Garcia, O. Schreiner, and W. Minker, 'A Scalable Syllable Speech Recognizer', in *Fortschritte der Akustik – DAGA '06*, edited by S. Langer, W. Scholl, and V. Wittstock (DEGA, Berlin, 2006), pp. 735–736.

[5] M. Tress, O. Schreiner, and G. Palm, 'Optimierung eines silbenbasierten Spracherkenners', in *Fortschritte der Akustik – DAGA '06*, edited by S. Langer, W. Scholl, and V. Wittstock (DEGA, Berlin, 2006), pp. 741–742.

[6] O. Schreiner, 'Modulation Spectrum for Pitch and Speech Pause Detection', in *Proc. 8th European Conference on Speech Communication and Technology (Eurospeech/Interspeech 2003)* (International Speech Communication Association, 2003), pp. 2849–2852.

[7] H. Quast, 'Automatische Erkennung nonverbaler Sprache', in *Fortschritte der Akustik – DAGA 2001*, edited by O. von Estorff (DEGA, Oldenburg, 2001), pp. 564–565.

[8] H. Quast, *Prosody Recognition in Speech Dialogue Systems* (Sierke Verlag, Göttingen, 2006), Doctoral thesis, Universität Göttingen.

[9] H. W. Strube, D. Michaelis, J. Lessing, and S. Anderson, 'Akustische Analyse pathologischer Stimmen in fortlaufender Sprache', in *Fortschritte der Akustik – DAGA '03*, edited by M. Vorländer (DEGA, Oldenburg, 2003), pp. 760–761.

[10] H. W. Strube, 'Sprach- und Bildanalyse für pathologische Stimmen', in *Signaltheorie und Signalverarbeitung, Akustik und Sprachakustik, Informationstechnik*, edited by D. Wolf (w.e.b. Universitätsverlag, Dresden, 2003), vol. 29 of *Studientexte zur Sprachkommunikation*, pp. 133–140.

[11] M. Fröhlich, D. Michaelis, and E. Kruse, 'Objektive Beschreibung der Stimmgüte unter Verwendung des Heiserkeits-Diagramms', HNO **46**, 684 (1998).

[12] D. Michaelis, M. Fröhlich, and H. W. Strube, 'Selection and combination of acoustic features for the description of pathologic voices', J. Acoust. Soc. Am. **103**, 1628 (1998).

[13] D. Michaelis, *Das Göttinger Heiserkeits-Diagramm – Entwicklung und Prüfung eines akustischen Verfahrens zur objektiven Stimmgütebeurteilung pathologischer Stimmen*, Doctoral thesis, Universität Göttingen, downloadable at http://webdoc.sub.gwdg.de/diss/2000/michaelis/ (2000).

[14] M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse, 'Acoustic Voice Analysis by Means of the Hoarseness Diagram', J. Speech, Language, and Hearing Res. **43**, 706 (2000).

[15] H. Kasuya, Y. Endo, and S. Saliu, 'Novel acoustic measurements of jitter and shimmer characteristics from pathological voice', in *Eurospeech '93 Proceedings*, edited by K. Fellbaum (Berlin, 1993), vol. 3, pp. 1973–1976.

[16] Y. Medan, E. Yair, and D. Chazan, 'Super resolution pitch determination of speech signals', IEEE Trans. Signal Processing **39**, 40 (1991).

[17] I. R. Titze and H. Liang, 'Comparison of F0 extraction methods for high precision voice perturbation measurement', J. Speech Hear. Res. **36**, 1120 (1993).

[18] H. Kasuya, S. Ogawa, Y. Kikuchi, and S. Ebihara, 'Normalized noise energy as an acoustic measure to evaluate pathologic voice', J. Acoust. Soc. Am. **80**, 1329 (1986).

[19] G. de Krom, 'A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals', J. Speech Hear. Res. **36**, 224 (1993).

[20] D. Michaelis, T. Gramss, and H. W. Strube, 'Glottal-to-Noise Excitation Ratio – a New Measure for Describing Pathological Voices', Acustica / acta acustica **83**, 700 (1997).

[21] T. Kohonen, *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences* (Springer, Berlin, 2001), 3rd ed.

[22] J. Kiosses and H. W. Strube, 'Akustische Charakteristiken der Normalstimme', in *Fortschritte der Akustik – DAGA 2001*, edited by O. von Estorff (DEGA, Oldenburg, 2001), pp. 84–85.

[23] M. Fröhlich, D. Michaelis, J. Lessing, and H. W. Strube, '"Breathiness measures" in acoustic voice analysis', in *Advances in Quantitative Laryngoscopy, Voice and Speech Research, Proc. 4th International Workshop*, edited by T. Braunschweig, J. Hanson, P. Schelhorn-Neise, and H. Witte (Friedrich-Schiller-Universität, Jena, 2000), pp. 63–71.

[24] A. van Ooyen and B. Nienhuis, 'Improving the Convergence of the Backpropagation Algorithm', Neural Networks **5**, 465 (1992).

[25] J. Lessing, *Entwicklung einer Klassifikationsmethode zur akustischen Analyse fortlaufender Sprache unterschiedlicher Stimmgüte mittels Neuronaler Netze und deren Anwendung*, Doctoral thesis, Universität Göttingen (2007).

[26] J. Lessing, M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse, 'Verwendung neuronaler Netze zur Stimmgütebeschreibung pathologischer Stimmen', in *Aktuelle phoniatrisch-pädaudiologische Aspekte 1998*, edited by M. Gross (Median, Heidelberg, 1999), vol. 6, pp. 39–43.

[27] J. Lessing, M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse, 'Akustische Stimmanalyse aus fortlaufender Sprache – Untersuchung von Tumorgruppen', in *Aktuelle phoniatrisch-pädaudiologische Aspekte 1998*, edited by M. Gross (Median, Heidelberg, 1999), vol. 6, pp. 126–130.

[28] J. Lessing, M. Fröhlich, D. Michaelis, H. W. Strube, and E. Kruse, 'A Neural Network Based Method to Assess Voice Quality from Continuous Speech for Different Voice Disorders', in *Advances in Quantitative Laryngoscopy, Voice and Speech Research, Proc. 4th International Workshop*, edited by T. Braunschweig, J. Hanson, P. Schelhorn-Neise, and H. Witte (Friedrich-Schiller-Universität, Jena, 2000), pp. 124–131.

[29] G. Fant, J. Liljencrants, and Q. Lin, *A four-parameter model of glottal flow*, Speech Transmission Laboratory – Quarterly Progress and Status Report 4/1985, Stockholm

(1985).
[30] M. Fröhlich, *Simultane Inversfilterung und Schätzung des glottalen Flusses aus akustischen Stimmsignalen*, Doctoral thesis, Universität Göttingen, downloadable at http://webdoc.sub.gwdg.de/diss/1999/froehlich/ (1999).
[31] M. Fröhlich, D. Michaelis, and H. W. Strube, 'SIM – simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signal', J. Acoust. Soc. Am. **110**, 479 (2001).
[32] A. El-Jaroudi and J. Makhoul, 'Discrete all-pole modeling', IEEE Trans. Signal Processing **29**, 411 (1991).
[33] S. Anderson, *Messung der Glottisöffnungsfläche und deren Beziehung zum Rauschanteil der Stimme*, Diploma thesis, Universität Göttingen, downloadable at http://sven.anderson.de/misc/diplomarbeit.pdf (2003).
[34] S. Anderson, D. Michaelis, and H. W. Strube, 'Vollautomatische Glottisdetektion bei Hochgeschwindigkeitsaufnahmen', in *Fortschritte der Akustik – DAGA '02*, edited by U. Jekosch (DEGA, Oldenburg, 2002), pp. 624–625.
[35] N. Nikolaidis and I. Pitas, *3-D Image Processing Algorithms* (John Wiley & Sons, 2001).

Oscillations, Waves and Interactions, pp. 37–71
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-03-7

On the use of specific signal types in hearing research

A. Kohlrausch[1,2] **and S. van de Par**[1]

[1]Digital Signal Processing Group, Philips Research Europe,
Eindhoven, The Netherlands
[2]Human-Technology Interaction, Eindhoven University of Technology,
Eindhoven, The Netherlands
Email: [1]*armin.kohlrausch@philips.com,* [2]*steven.van.de.par@philips.com*

Abstract. In this contribution, we review a number of specific signal types that have been introduced in auditory research in the past 20 years. Through the introduction of digital computers into experimental and theoretical hearing research, the freedom to construct and use specific acoustic stimuli in behavorial and also physiological research has grown steadily. In parallel, the use of computer models allowed to analyze and predict, within certain limits, how specific properties of acoustic stimuli influence the perception of a listener. As in other fields of physics, the close interplay between experimental tests and quantitative models has been shown to be essential in advancing our understanding of human hearing.

1 Introduction

One of the scientific areas to which the research groups at the Dritte Physikalische Institut (DPI) contributed significantly is the wide field of psychoacoustics. The interest in this area can be traced back to early research activities of its first director, Erwin Meyer [1]. Evidence of his strong and continuing interest is provided by the fact that he chose the opportunity of his inaugural lecture after his new appointment at the university of Göttingen, which he gave in January 1948, to talk about: 'Über den derzeitigen Stand der Theorie des Hörens (On the current state of the theory of hearing)' [2].

One of the best-known psychoacoustic contributions coming from the DPI is based on the work by Haas [3]. The Haas-effect refers to the observation that a reflection that follows the direct sound with a short delay up to 50 ms can be significantly higher in level than the direct sound, without being perceived as annoying. In later years, hearing-related problems were mostly studied in the context of room-acoustic questions, and most perceptual activities in the 1970's were devoted to the broad area of subjective room acoustics [4].

The classical psychoacoustic studies, based on well-controlled acoustic stimuli delivered via headphones to subjects sitting in a sound-isolated, and often very narrow,

listening booth, increased strongly throughout the 1980's. This increased interest was reflected even in a rebuilding of the central space in the "Halle" of the DPI, in front of the "Reflexionsarmer Raum", where the control panels for the loudspeaker dome were dismantled and spaces for two listening booths were created. In parallel with the acoustic spaces, also the computer infrastructure for controlling listening experiments and generating signals with more complexity than contained in Gaussian noise or sinusoids grew steadily in this period. In our view, this experimental infrastructure together with a growing group of young scientists were essential for the increasing level of sophistication of hearing research at the DPI. In the following, we want to describe and analyze one of the factors, namely the creativity in using signals with specific spectral and temporal properties in listening experiments and model simulations. This creativity started at the DPI, but was spread to other places like Eindhoven and Oldenburg, and more recently to Lyngby/Kopenhagen, and has influenced many research paradigms in hearing research groups all over the world.

2 Harmonic complex tone stimuli

The first class of signals we will discuss are signals with a periodic waveform. Depending on the way of construction of those signals, they are either considered in the context of their temporal properties (e. g., when a regular series of clicks is generated and the perceptual influence of a slight temporal deviation from regularity is considered, see, e. g., Ref. [5]) or in the context of their spectral properties (consider, e. g., the role of the vocal tract filter on the resulting vowel quality). Of course, from a mathematical point of view, time-domain and spectral descriptions are fully equivalent, if indeed not only the power spectrum, but the full complex spectrum, including the phase, is considered. Historically, however, temporal and spectral views were quite distinct, mainly under the influence of signal analysis systems that allowed to represent the power spectrum, but not the phase spectrum. Also the paradigm of critical bands and auditory filters, which for a long time were only defined in terms of their overall bandwidth and their amplitude characteristics (see, e. g., Ref. [6]), made it difficult to bring the temporal and spectral views closer together. Again, the increasing use of computer programs to generate acoustic stimuli and to perform time-domain modeling of perceptual processes emphasized the role of the phase spectrum on the perceptual quality of periodic signals [7]. In the following, we will focus on the description of one specific type of complex tones, those with so-called Schroeder phases.

2.1 Schroeder-phase harmonic complex tones

2.1.1 Definition

The term Schroeder phase refers to a short paper by Schroeder from 1970 [8]. In this paper, he addressed the problem how the peak-to-peak amplitude of a periodic waveform with a given power spectrum can be minimized. He related this problem to the observation that frequency-modulated stimuli have a low peak factor. The proposed solution lies in a phase choice which gives the signal an FM-like property.

The general solution derives the individual phase values without any restriction for the power spectrum. A more specific case, which has found its way into hearing research, is that of a harmonic complex with a flat power spectrum. The solution for the individual phase values of such a complex is as follows:

$$\phi_n = -\pi n(n-1)/N\,, \tag{1}$$

with N the total number of components in the complex. The important term in this equation is the quadratic relation between component number, n, and component phase, ϕ_n, which will induce an approximately linear increase in instantaneous frequency. The normalization with N creates a signal, for which the instantaneous frequency sweeps once per period from the frequency of the lowest to that of the highest component in the complex. In fact, the instantaneous frequency has a periodic sawtooth-like course for such signals.

It is obvious that reversing the initial sign in Eq. (1) has no influence on the peak factor of the resulting signal, but it will invert the direction of the linear frequency sweep. Because these two versions of a Schroeder-phase signal lead to substantially different percepts, a convention has been introduced to distinguish them. A *negative* Schroeder-phase signal is a signal were the phases of individual components are chosen as in Eq. (1). In contrast, a *positive* Schroeder-phase signal has phase values with a positive sign in front of the fraction. One can memorize this relation by using the fact that the sign of the phase is opposite to the change in instantaneous frequency.

2.1.2 Acoustic properties

By construction, Schroeder-phase stimuli have a relatively flat temporal envelope and the peak factor, defined as the relation between envelope maximum relative to the rms-value of the signal, is much lower than for other phase choices. This is demonstrated in Fig. 1, in which the waveforms for harmonic complexes composed of 19 equal-amplitude harmonics are compared for three different choices of the component phase: Besides positive and negative Schroeder-phase stimuli, the top part shows the waveform of a zero-phase stimulus, all with a spectrum from 200 Hz to 2000 Hz. For this latter stimulus, the energy is concentrated at very short instances within each period, leading to a much higher peak factor. The spectro-temporal properties of these signals can be seen more clearly in a short-time spectral representation. Figure 2 shows the spectra of the three signals from Fig. 1 calculated using a moving 5-ms Hanning window. The sawtooth-like frequency modulation of the two Schroeder-phase complexes is pronounced in this representation. In addition, the plot for the zero-phase complex shows ridges at the spectral edges of 200 and 2000 Hz. These relative spectral maxima can be perceived as pitch, superimposed on the 100-Hz virtual pitch of the complexes, and the presence of these pitch percepts has been used as a measure of the internal representation of such harmonic complexes [10,11]. The visibility of the temporal structures in the short-time spectrum depends critically on the duration of the temporal window, relative to the period of the sound. The shorter the window, the stronger temporal changes are visible, the longer the window, the stronger the power spectral properties are emphasized.

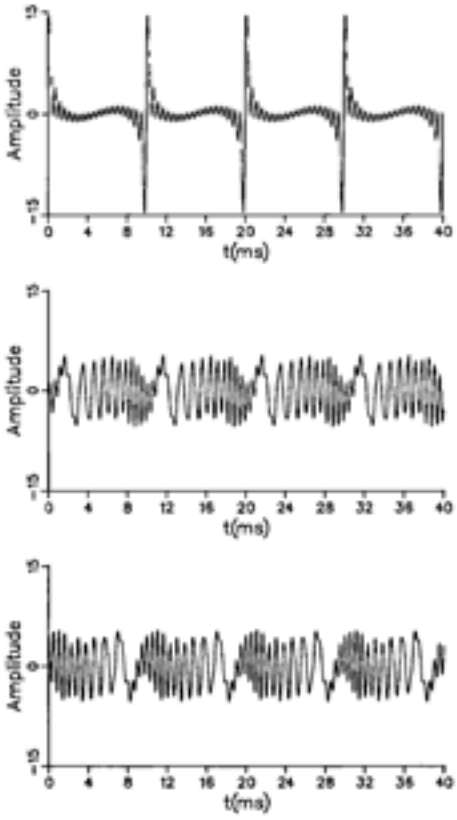


Figure 1. Time functions of harmonic complexes for three different choices of the component's starting phases. Top: $\phi_n = 0$ zero-phase complex, middle: $\phi_n = -\pi n(n-1)/N$, negative Schroeder-phase complex, bottom: $\phi_n = +\pi n(n-1)/N$, positive Schroeder-phase complex. All complexes are composed of the equal-amplitude harmonics 2 to 20 of fundamental frequency 100 Hz. For this plot, the amplitude of an individual harmonic was set at 1. Reused with permission from Ref. [9]. Copyright 1995, Acoustical Society of America.

2.1.3 Role in hearing research and perceptual insights

The first paper in which the Schroeder-phase formula was used in hearing experiments was published by Mehrgardt and Schroeder in the proceedings of the 6th International Symposium on Hearing, 1983 [12]. In this paper, the quadratic phase formula from Eq. (1) was combined with an additional scaling factor, which allowed to control the spread of signal energy throughout the period. The spectrum of the harmonic complex was, however, not flat as in most later investigations, but the individual components had Hanning-weighted amplitudes. This paper emphasized the influence of the masker's temporal waveform on the observed masking behavior and showed, how strongly the acoustic waveform can vary by just varying the phase spectrum.

The great potential of Schroeder-phase signals to observe the phase characteristic of the auditory filter was found out quite accidentically. During his master thesis research, Bennett Smith was interested in acoustic figure-ground phenomena, where spatial orientation in the visual domain was translated into linear frequency modulation in the auditory domain [13]. In the construction of his acoustic background stimuli, he made use of the Schroeder-phase formula. He did, unfortunately, not find any effect of acoustic figure-ground orientation on audibility, but made instead an-

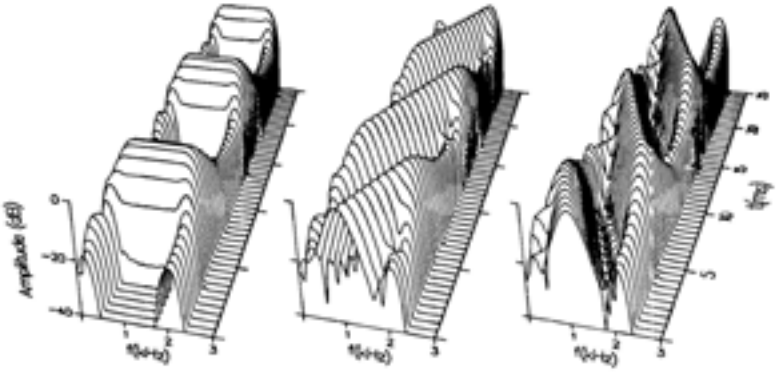


Figure 2. Short-time spectral representation of the signals from Fig. 1 using a Hanning window with length 5 ms. The left panel shows the zero-phase signal, the middle panel the negative Schroeder phase and the right panel the positive Schroeder phase complex. Reused with permission from Ref. [9]. Copyright 1995, Acoustical Society of America.

other observation: the audibility of a sinusoidal stimulus in a complex-tone masker depended strongly on the sign in the phase formula: When the background was constructed with a positive sign, the masked thresholds were lower by up to 20 dB, compared to the situation in which the phase sign was negative.

This large threshold difference formed a considerable scientific puzzle: both masker versions had a similarly flat temporal envelope, so there was no reason to assume a difference in masking potential for simultaneously presented sinusoids. The effect was finally understood on the basis of computer simulations with a time-domain basilar-membrane model which had been realized by Hans-Werner Strube [14]. This model was, at that time, the first computer model with a realistic phase characteristic for the basilar membrane, which allowed to compute the basilar-membrane output in the time domain, and it was sufficiently time-efficient. Strube observed that the two versions of the Schroeder-phase complex lead to very different waveforms at the output of the filter element that represented a specific place on the basilar membrane. One waveform was clearly more modulated, and the existence of valleys in the masker waveform fitted nicely to the observation by Smith that this masker also led to lower masked thresholds: the target signal could be detected easier in these masker valleys.

In the next years, a great number of further experiments and model simulations were performed [9,15,16] which led to the following insights. The clue to understand the differences between positive and negative Schroeder-phase stimuli lies in the phase characteristic of the auditory filter. If subjects have to detect a narrowband stimulus in a broadband masker, only the frequency region around the target frequency is of interest. In order to simulate the transformations in the auditory periphery for such an experiment, we have to compute the waveforms of the acoustic stimuli at the output of the auditory filter that is centered on the target frequency. This waveform will depend critically on the amplitude and the *phase* characteristic of the auditory filter.

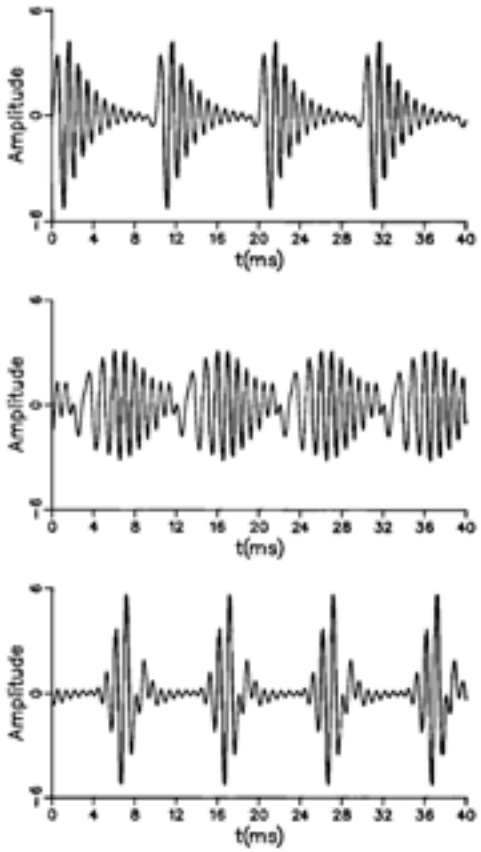


Figure 3. Responses of a linear basilar-membrane model at resonance frequency 1100 Hz to the three harmonic maskers with fundamental frequency 100 Hz shown in Fig. 1. The top panel shows the zero-phase signal, the middle panel the negative Schroeder phase and the bottom panel the positive Schroeder phase complex. Reused with permission from Ref. [9]. Copyright 1995, Acoustical Society of America.

The most important conclusion was that, for the right choice of input parameters, the phase characteristic of a Schroeder-phase stimulus matches quite closely this phase characteristic of the auditory filter, at least in the spectral region of maximum transfer of the filter. Since Schroeder-phase stimuli come in two flavours, one version, the negative Schroeder-phase stimulus, will have the same phase characteristic as the auditory filter while the positive Schroeder-phase stimulus has a phase spectrum that is *opposite* to that of the filter. In the transfer through a filter, the input phase spectrum and the filter phase spectrum add to give the phase spectrum of the output signal. For the positive Schroeder phase, we thus have the situation of phase compensation, and the resulting filtered signal at the output of the auditory filter has a nearly constant phase of all components. Conceptually, this situation is quite similar to pulse compression through frequency modulation, as it is used in radar and sonar technology. In a way, the positive Schroeder-phase stimuli are matched in their phase characteristic to the auditory filters as they are realized mechanically in the inner ear.

This interpretation has some interesting consequences: if a specific Schroeder-phase stimulus is optimally matched in its phase to the inner ear filter at a certain

frequency, then such a signal should have a higher peak factor after filtering than a zero-phase input stimulus. This relation is analyzed in Fig. 3, which shows waveforms at the output of a linear basilar-membrane filter tuned to 1100 Hz. The two panels at the bottom show the waveforms of the two Schroeder-phase complexes. While the broadband input signals to the filter have a flat temporal envelope (cf. Fig. 1), both waveforms have a clear amplitude modulation after filtering, which follows the 10-ms periodicity of the stimulus. But the depth of modulation is quite different for the two signals, the positive Schroeder-phase stimulus at the bottom has a much higher peak factor than the negative complex in the middle. This simulation reflects the initial observation made by Strube in Ref. [14].

The top panel shows the filtered version of the zero-phase complex, which has a highly peaked input waveform. The filtered waveform shows, within each period, the impulse response of the auditory filter, because the zero-phase complex is similar to a periodic sequence of pulses. Comparing the top and the bottom panel reveals that, indeed, the positive Schroeder-phase complex has a somewhat higher peak factor than the zero-phase complex, and its energy is more concentrated in time, as expected based on the pulse-compression analogon.

The relation between zero-phase and positive-Schroeder-phase stimuli formed also the key to estimate the phase properties of a specific point on the basilar membrane. If we were able to determine the phase curvature, for which the match between stimulus and filter phase is "optimal", then this value was an indication for the phase curvature (or the frequency-dependent group delay) of the filter. The clue to such an analysis is given by comparing perceptual thresholds for positive Schroeder-phase maskers with those for zero-phase maskers. For maskers, for which this difference is largest (and for which the Schroeder-phase stimulus as masker gives lower masked thresholds), the phase curvature at the signal frequency is an estimate of the filter phase. Figure 4 shows data from Ref. [9] which were used for such a computation.

In the region of f_0 values between 100 and 150 Hz, the thresholds obtained with the positive Schroeder-phase complex (left-pointing triangles) and with the zero-phase complex (squares) show the largest difference. For theses complexes, the second derivative of the phase-versus-frequency relation, which indicates the phase curvature has values between $1.05 \times 10^{-5}\pi/\mathrm{Hz}^2$ and $0.74 \times 10^{-5}\pi/\mathrm{Hz}^2$. We can conclude that the curvature of the phase characteristic for the basilar-membrane filter centered at 1100 Hz should be in the range of these two values. A similar conclusion about the phase curvature can be derived from the parameters of those complexes, for which positive Schroeder-phase and zero-phase complexes lead to approximately the same threshold. In this case, the internal envelope modulation of the two complexes after filtering on the basilar membrane should be approximately equal. As explained in detail in Ref. [9], the phase curvature in the Schroeder-phase stimuli should be half the value of the filter curvature, and this is reached for fundamental frequencies of 50 to 75 Hz. And exactly in this region, the two lower curves in Fig. 4 cross each other.

This consideration allowed a first computation of the auditory filter phase for one frequency, 1100 Hz. In Ref. [9], additional threshold measurements were included for frequencies 550, 2200 and 4400 Hz, thus covering a range of three octaves. It is often assumed that the auditory filter has a nearly constant quality factor across the range of audible frequencies. If this was also true for the phase characteristic, then the

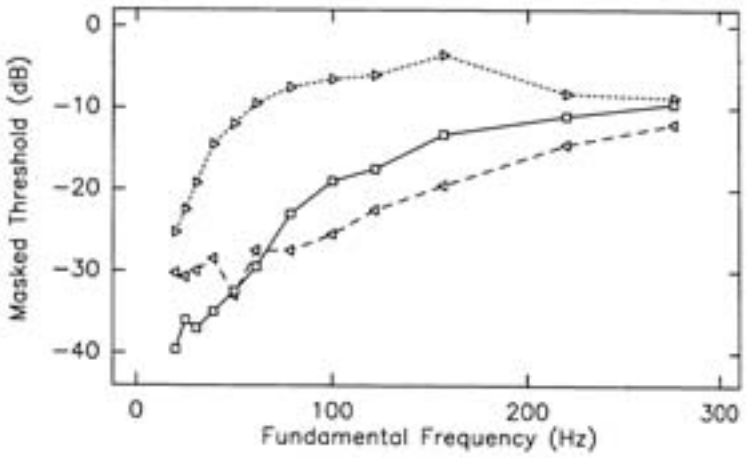


Figure 4. Simultaneous masked thresholds of a 260-ms, 1100-Hz signal as a function of the fundamental frequency f_0 of the harmonic complex masker. Thresholds are expressed relative to the level of a single masker component. The maskers were presented at a level of 75 dB SPL. Squares: zero-phase complex; right-pointing triangles: negative Schroeder-phase complex; left-pointing triangles: positive Schroeder-phase complex. Reused with permission from Ref. [9]. Copyright 1995, Acoustical Society of America.

results obtained at 1100 Hz would allow a direct prediction for the phase characteristic in the range of three octaves around 1100 Hz. The comparison with the results at 550 Hz indeed revealed the expected relation, while towards higher frequencies, the curvature changed somewhat less with center frequency than expected for a system, in which the amplitude and phase characteristics of the filters remain constant on a logarithmic frequency scale.

One final important observation from these initial Schroeder-phase studies needs to be mentioned. The view on the shape of the auditory filters was in the 1980's strongly influenced by the work of Patterson, Moore and colleagues, who had used the notched-noise technique to estimate the amplitude characteristic of the auditory filter. The best characterization was possible with a so-called rounded exponential filter shape [17]. A time-domain implementation of a filter with such an amplitude characteric was possible based on so-called gamma-tone filters [18,19]. Due to the large amount of studies supporting this concept of auditory filters, we were interested to analyze the Schroeder-phase stimuli with such a filter.

Figure 5 presents, in a similar format as Figs. 1 and 3, four periods of the waveform for harmonic complexes with fundamental frequency 100 Hz. The analysis shows the output of the gamma-tone filter tuned to 1100 Hz, and the three subpanels are for the three different phase choices. It is apparant that this filter does not lead to differences in the modulation depth between the three stimuli, and based on this simulation one would expect quite similar masking behaviour of all three complex tones, in contrast to the experimental data. The major reason for the similar treatment of the two Schroeder-phase maskers by the gamma-tone filter is its antisymmetric phase

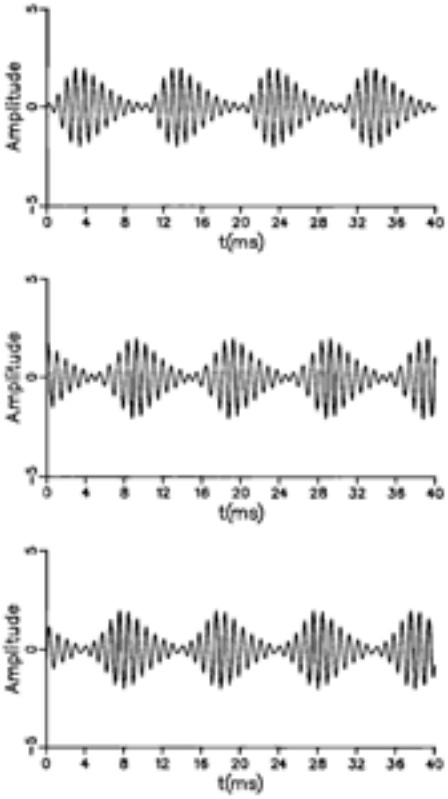


Figure 5. Responses of a linear, fourth-order gamma-tone model at resonance frequency 1100 Hz to the three harmonic maskers with fundamental frequency 100 Hz shown in Fig. 1. The top panel shows the zero-phase signal, the middle panel the negative Schroeder-phase and the bottom panel the positive Schroeder-phase complex. Reused with permission from Ref. [9]. Copyright 1995, Acoustical Society of America.

characteristic relative to its resonance frequency. The curvature of the filter phase changes its sign at the resonance frequency from negative to positive. A filter with such a phase characteristic can never flatten out the phase of a Schroeder-phase complex over the full range of its passband.

2.1.4 Later developments

Although the first paper on Schroeder-phase stimuli was already published in 1986 [15], the paradigm was only widely adopted after publication of our second paper in 1995 [9]. The first papers that used the term "Schroeder phase" in their title were published in 1997 [20,21]. Many authors related psychoacoustic findings with Schroeder-phase stimuli to the properties of the basilar membrane. Differences that were found between normal-hearing and hearing-impaired subjects and also influences of the overall presentation level indicated some role of active processes in creating large differences between positive and negative Schroeder-phase stimuli [21–23]. Based on the results of these studies, Summers concluded: "The current results showed large differences in the effectiveness of positive and negative Schroeder-phase maskers under test conditions associated with nonlinear cochlear processing. The two maskers were more nearly equal in effectiveness for conditions associated with more linear pro-

cessing (high levels, hearing-impaired listeners). A number of factors linked to the cochlear amplifier, including possible suppressive effects and level-dependent changes in the phase and magnitude response of effective filtering, may have contributed to these differences." (Ref. [23], p. 2316).

The analysis of the phase characteristics of auditory filters was further refined by Oxenham and Dau [24,25]. They varied the phase curvature of Schroeder-phase complexes by using a scalar multiplier in front of Eq. (1), very similar to the use of the Schroeder-phase formula by Mehrgardt and Schroeder [12]. They concluded that the scaling invariance of filter phase with filter center-frequency, as expected for a set of filters with constant quality factor, might hold for frequencies above 1 kHz, but not for lower frequencies.

Schroeder-phase stimuli have also been used in physiological experiments, which allowed a direct test of the basic hypothesis of the role of peripheral filtering on modulation depth of the waveform, as published in 1985 by Strube. In 2000, Recio and Rhode [26] measured the basilar membrane response in the chinchilla for positive and negative Schroeder-phase stimuli and also for clicks, thus using very similar types of stimuli as the early psychoacoustic studies. They concluded: "The behavior of BM responses to positive and negative Schroeder complexes is consistent with the theoretical analysis performed by Kohlrausch and Sander in 1995, in which the curvature i. e., the second derivative of the phase versus frequency curve of the BM was used to account for the differences in the response to each of the two Schroeder phases. [...] Hence, phase characteristics of basilar membrane responses to positive Schroeder-phase stimuli show reduced curvatures (relative to the stimulus), and, as a result, peaked waveforms (Kohlrausch and Sander, 1995)" (Ref. [26], p. 2296).

3 Noise signals with non-Gaussian statistic

An important class of signals used in hearing science are noise signals. Although the meaning of the term noise is wider in daily use, in hearing sciences, it refers to signals that are inherently random. When we consider for example white Gaussian noise, samples taken from its temporal waveform are randomly distributed according to a Gaussian distribution and samples taken at subsequent moments in time are uncorrelated. The frequency domain representation of white Gaussian noise shows a complex spectrum where the real and imaginary parts are also Gaussian distributed. The spectrum is called white because the signal energy is uniformly distributed across frequency.

Often noise signals are subjected to some kind of spectral filtering. Although this influences the spectral envelope of the signal, the Gaussian distribution of the time domain samples and the complex spectral components are not influenced. The correlation, however, across samples taken at different moments in time is influenced. This is reflected in the autocorrelation function. For a white noise signal, the autocorrelation function is peaked at lag zero, and zero at all other lags, in line with the idea that samples are mutually uncorrelated. For a filtered noise, however, there will be correlations across samples which is reflected in the autocorrelation function at lags different than zero.

Noise signals have been extensively used to study auditory masking where they often serve as masker signals. For example, in the early experiments related to critical bandwidth by Hawkins and Stevens [27], white Gaussian noise maskers were used to measure the dependence of masked thresholds of a tonal signal as a function of frequency. Probably the preference for using noise signals as a masker is related to its uniform energy distribution across time and frequency and the fact that it is a well defined signal. The inherently stochastic nature of noise, however, has implications for its masking behavior as was demonstrated in studies that employed reproducible noise for which the stochastic uncertainties in the noise are effectively removed. Generally, reproducible noise produces lower masked thresholds than running noise [28].

3.1 Low-noise noise

3.1.1 Definition

In the previous section we discussed that a filtering operation on white Gaussian noise causes the autocorrelation function to change from a delta function at lag zero, to a pattern that reflects predictability of successive time samples of such noise. This predictability is reflected in a smooth development of the envelope of the time domain noise waveform[1]. The rate of fluctuation in the temporal envelope is proportional to the bandwidth of the filtered noise signal. The spectrum of the envelope has a large DC component and a downward tilting slope that leaves very little spectral power beyond frequencies equal to the bandwidth of the bandpass noise. Interestingly, the *degree* of fluctuation is independent of the bandwidth, which is reflected in the probability density function of the temporal envelope values which is Rayleigh distributed. Thus, there is an inherently high degree of fluctuation in Gaussian noise.

The inherent fluctuations that are present in Gaussian noise have prompted the development of so-called low-noise noise [29]. This special type of noise has the same spectral envelope as Gaussian noise, but a much lower degree of inherent fluctuations in its temporal envelope, hence the name *low-noise* noise. This allowed the study of the contribution of envelope fluctuations to auditory masking phenomena by comparing the masking effect of Gaussian and low-noise noise. The first to pursue this idea where Pumplin and Hartmann [30].

3.1.2 Stimulus generation

The original manner to generate low-noise, such as promoted by Ref. [29] was via a special optimization algorithm. First a band-pass noise was digitally generated in the frequency domain by setting amplitudes in a restricted spectral range to some specific values, e. g., one constant value, and randomizing the phases. Such a noise will approximate all the properties of a bandpass Gaussian noise when the product of time and frequency is sufficiently large. Via a steepest descent algorithm, the phase spectrum was modified step-by-step in the direction which made the temporal enve-

[1] There are alternative manners to determine the envelope of a signal which lead to somewhat different envelopes. We will consider here the Hilbert envelope.

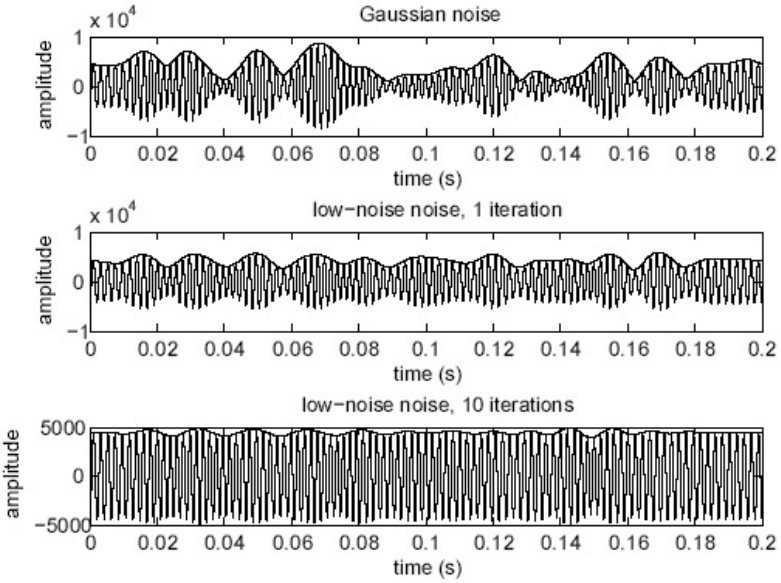


Figure 6. Illustration of the low-noise noise generation. The top panel shows the time-domain Gaussian noise at the start of the iterative process, the middle panel the low-noise noise after one iteration, the lower panel, the low-noise noise after 10 iterations. All waveforms are shown with their respective envelopes.

lope more flat, according to some statistical measure[2]. After a sequence of iterations, a low-noise noise waveform was obtained with a rather flat temporal envelope and the initial amplitude spectrum. Thus, summarizing, the method of Pumplin obtained low-noise noise by modifying the phase spectrum in a special way.

Later on, several alternative manners to generate low-noise noise were proposed and evaluated by Kohlrausch et al. [31]. We will here describe the method that led to the lowest degree of fluctuation in the temporal envelope. The method consists of an iterative process that is initiated by generating a time-discrete Gaussian bandpass noise. The iterative process then consists of a sequence of straightforward steps.

First the Hilbert envelope of the noise is calculated. Secondly, the noise waveform is divided by its Hilbert envelope on a sample-to-sample basis in the time domain. For the rare occasions that the Hilbert envelope is equal to zero, the resulting division is set to zero. In the third step, a bandpass filtering is applied to remove the new spectral components outside of the specified bandpass range that were introduced by the division operation in the previous step. By repeating the iterative steps several times, a much flatter envelope is obtained.

After the first two steps, calculating the Hilbert envelope and dividing the noise waveform by its Hilbert envelope, the resulting temporal waveform will have a flat envelope. The spectrum will also be modified considerably. The division by the Hil-

[2]normalized fourth moment of the temporal envelope distribution

Number of iterations	normalized fourth moment
0	3.030
1	1.845
2	1.701
4	1.591
6	1.552
8	1.535
10	1.526
Hartmann & Pumplin	1.580

Table 1. Normalized fourth moment of low-noise noise as a function of the number of iterations. The bottom row gives the value from Hartmann and Pumplin [30] for comparison.

bert envelope can be seen as a time-domain multiplication with the reciprocal Hilbert envelope. In the frequency domain, this is equivalent to a convolution of the bandpass noise signal with the spectrum of the reciprocal Hilbert envelope. Due to the large DC component present in the envelope, also the reciprocal envelope will have a large DC component. Thus, the convolution in the frequency domain will be dominated by this DC component and as a consequence, the spectrum of the bandpass noise will remain largely intact. However, there will be additional, new spectral components that are outside the bandpass range of the original bandpass noise.

Therefore, in the third step, bandpass filtering is applied to remove the new spectral components outside of the specified bandpass range that were introduced by the division operation in the previous step. Considering the argumentation given above, only a relatively small amount of signal power is removed by this operation. Nevertheless, the temporal envelope will not be flat anymore. In Fig. 6, the temporal waveforms are shown for the original 100-Hz wide Gaussian bandpass noise centered at 500 Hz, that was input to the iterative process (top panel), and after the first iteration of our algorithm (middle panel). As can be seen, the degree of envelope fluctuation is reduced considerably. By repeating the iterative steps several times, a much flatter envelope is obtained after 10 iterations (lower panel). Convergence is assumed to be obtained due to the DC component in the Hilbert envelope becoming more dominant over the higher spectral components after each iteration.

In Fig. 7, the same signals are shown, only now represented in the frequency domain. As can be seen, the original, bandpass Gaussian noise has a uniform spectral envelope. The spectrum of the low-noise noise signal is, even after 10 iterations, quite similar to the spectrum of the Gaussian signal. There is, however, a tendency for the spectrum to have a somewhat lower level towards the edges of the bandpass range.

As a measure of envelope fluctuation, Table 1 shows the normalized fourth moment for different numbers of iterations of our algorithm. The value obtained by Hartmann and Pumplin [30] is shown at the bottom of the table. As can be seen, already after 6 iterations, we obtain a lower degree of envelope fluctuation than the method of Hartmann and Pumplin. After 10 iterations, the normalized fourth moment is 1.526, close to the theoretical minimum of 1.5 for a sinusoidal signal.

In summary, the iterative method is able to create a low-noise noise by modifying both the phase and the amplitude spectrum. The specific ordering of spectral com-

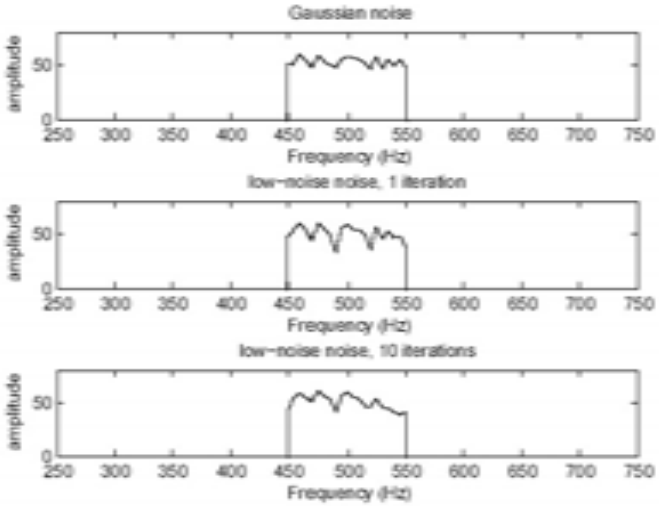


Figure 7. Illustration of the low-noise noise generation. The top panel shows the power spectrum of the Gaussian noise at the start of the iterative process, the middle panel the spectrum of the low-noise noise after one iteration, the lower panel the spectrum of the low-noise noise after 10 iterations.

ponents in the passband causes the flat envelope that is seen in the lower panel of Fig. 6. Due to this careful arrangement of phase and amplitude values throughout the noise spectrum, any modification of this spectral ordering will affect the flatness of the temporal envelope. In Fig. 8, the low-noise noise signal of Fig. 6, which was centered at 500 Hz, and had a bandwidth of 100 Hz, is shown after being filtered with a 78-Hz-wide gamma-tone filter centered at 500 Hz. As can be seen, the degree of envelope fluctuation has increased considerably. Since the gamma-tone filter used here is a reasonable first-order approximation of auditory peripheral filtering, this figure demonstrates that the properties that are present in the external stimulus should not be taken to be representative for the manner in which the stimulus is represented within the auditory system.

3.1.3 Role in hearing research and perceptual insights

As discussed before, Gaussian noise is a frequently used signal to serve as a masking stimulus in experiments investigating auditory masking. Early experiments of Fletcher [32] used noise signals of various bandwidths to determine detection thresholds of sinusoidal signals centered in the bandpass noise maskers. In these experiments it was found that only the masker energy that was spectrally close to the sinusoidal target signal contributed to the masking effect of the noise. This led to the concept of the critical band which indicates the spectral range that contributes to the masking effect on the sinusoidal signal. The integrated intensity of the masker within this range determines the masked threshold.

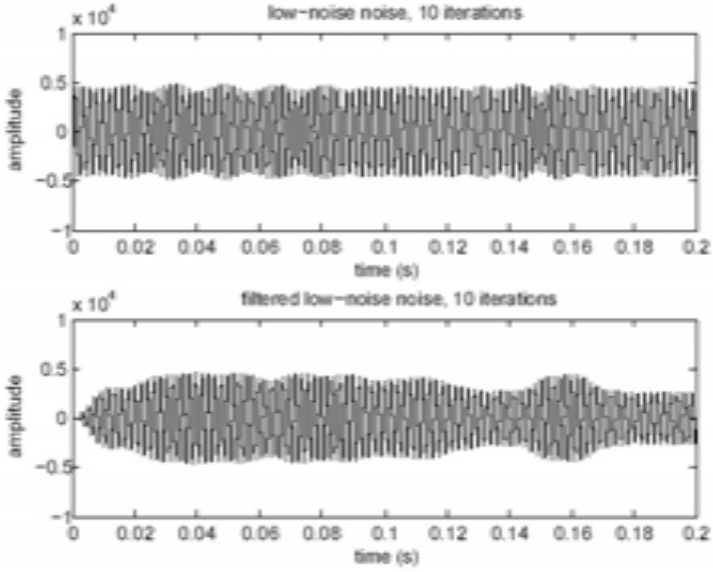


Figure 8. Illustration of the low-noise noise generation. The top panel shows a low-noise noise with a bandwidth of 100 Hz after 10 iterations, the lower panel shows the same waveform after peripheral filtering with a gamma-tone filter of 78-Hz width.

This purely intensity-based account of masking does not provide insights into the reasons for observing quite different masked thresholds when narrow-band noises or sinusoidal signals with the same overall level are used as a masker. When the bandwidth of the noise is smaller than the critical bandwidth, there is no difference in the masker intensity within that critical band and, if overall masker intensity determines masking, thresholds should be the same. Typically, however, thresholds for tonal maskers are about 20 dB lower than for narrowband Gaussian noise maskers [33].

One of the factors that is believed to contribute to the different masking strength of these signals is the difference in the inherent envelope fluctuations. A tonal masker has no inherent envelope fluctuations, and the addition of the target tone will introduce a beating pattern which may be an effective cue for detecting the presence of the target. In a noise masker, however, the masker itself already has a high degree of fluctuation. Addition of the sinusoidal signal does not alter the properties of the envelope fluctuations by a significant degree and therefore, changes in the temporal envelope pattern may be a less salient cue for a noise masker.

Low-noise noise maskers provide an elegant stimulus to verify that the inherent fluctuations in Gaussian noise are an important factor contributing to its strong masking effect. Such an experiment had been done by Hartmann and Pumplin [30], but the difference that they found was only 5 dB. This difference is considerably smaller than the 20-dB difference in masking found for Gaussian noise maskers and tonal signals. A complicating factor may be that the inherent fluctuations in the

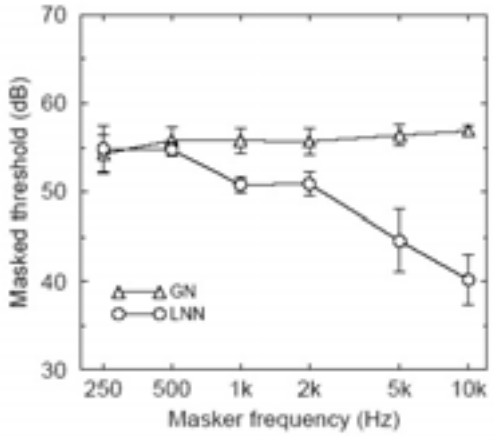


Figure 9. Masked thresholds for 100-Hz wide Gaussian-noise (triangles) and low-noise noise maskers (circles) as a function center frequency. Reused with permission from Ref. [31]. Copyright 1997, Hirzel Verlag and European Acoustics Association.

low-noise noise that was used by Pumplin and Hartmann were still strong enough to cause a significant masking effect. Furthermore, the bandwidth of their low-noise noise stimulus was 100 Hz around a center frequency of 500 Hz. Although such a bandwidth agrees approximately with the estimates of auditory filter bandwidth at this frequency, peripheral filtering may have caused a significant reduction of the flatness of their low-noise noise stimulus, as we demonstrated in Fig. 8, where a low-noise noise signal with their spectral properties was filtered with a 1-ERB wide filter.

A more recent experiment by Kohlrausch et al. [31] used low-noise noise created by the iterative method outlined in the previous section which results in an even lower degree of inherent fluctuation. In addition, the experiment of Kohlrausch et al. measured masked thresholds as a function of center frequency of the low-noise noise masker while keeping the target tone always spectrally centered within the 100-Hz wide noise masker. The highest center frequency in their experiment was 10 kHz, a frequency where the peripheral filter bandwidth is considerably larger than the masker bandwidth. As a result it can be assumed that the peripheral filtering will only have a marginal effect on the temporal envelope flatness of the low-noise noise. Thus in these experiments, the difference in masking thresholds between Gaussian noise and low-noise noise should be about the same size as the difference seen for Gaussian noise and tonal maskers.

The results of the experiments by Kohlrausch et al. [31] are shown in Fig. 9. As can be seen the masked thresholds for Gaussian noise (triangles) are constant for center frequencies of 500 Hz and above. This is in line with the fact that auditory filtering does not reduce masker intensity for a 100-Hz wide masker in this frequency range, and that the degree of inherent envelope fluctuations does not vary as a function of

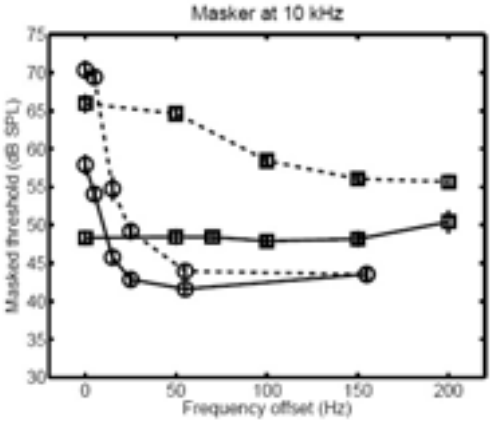


Figure 10. Masked thresholds for Gaussian noise (dashed lines) and low-noise noise maskers (solid lines) as a function of frequency offset between the masker and target. Squares show 100-Hz wide maskers and circles 10-Hz wide maskers both presented at 70 dB SPL.

center frequency. For low-noise noise (circles), however, we see a clear dependence of thresholds on the center frequency. Although low-noise noise thresholds were lower than Gaussian noise thresholds already at a center frequency of 1 kHz, for 10 kHz we see a much larger difference of more than 15 dB which is much more similar to the difference observed for sinusoidal and Gaussian noise maskers. The higher thresholds for lower frequencies are well in line with the idea that peripheral filtering affects the temporal envelope flatness of low-noise noise.

A variant of the experiment by Kohlrausch et al. [34] investigated the effect of a frequency offset between masker and sinusoidal target signal with the target always higher in frequency than the masker [35]. The addition of the sinusoidal target to the masker band creates modulations with a rate that is characterized by the frequency difference between target and masker. When the target is centered within the masker, the newly introduced modulations will have a rate comparable to those already present within the masker alone and will therefore be difficult to detect. When the target is sufficiently remote from the masker band, the modulations that will be introduced due to addition of target will be of considerably higher rate than those already present within the masker and may be much easier to detect. There is evidence that the auditory system exhibits some frequency selectivity associated with the processing of temporal envelope fluctuations, which led to the modulation filterbank model proposed by Dau et al. [36].

In Fig. 10 results of the experiment by van de Par and Kohlrausch [35] are shown. Thresholds for Gaussian noise (dashed lines) and low-noise noise (solid lines) maskers are shown centered at 10 kHz, for various target-to-masker frequency offsets and two masker bandwidths. As can be seen for the squares showing the 100-Hz wide maskers, the low-noise noise thresholds (squares with solid lines) are roughly independent of frequency offset. The Gaussian noise thresholds for the same bandwidth (squares

with dashed lines), however, show a clear dependence of frequency offset and thresholds are generally higher than the low-noise noise thresholds, in line with the idea that the inherent fluctuations in the Gaussian masker prohibit the detection of the modulations introduced through the addition of the sinusoidal target signal. At larger frequency offsets, however, the modulations introduced by the target signal become higher in rate, and thresholds are relatively low. Note that the frequency offsets are considerably lower than peripheral filter bandwidths. Thus the patterns of thresholds observed in Fig. 10 are not likely to be influenced by peripheral filtering. For the 10-Hz maskers (circles), also for the low-noise noise (solid lines) there is a dependence of frequency offset suggesting that there are inherent fluctuations in the low-noise noise masker that influence masking at small frequency offsets. Note that for small offsets the low-noise noise thresholds are considerably lower than the Gaussian noise thresholds (circles with dashed lines). Again the dependence on frequency offset that is observed here is a reflection of the processing of temporal envelope fluctuations and not of spectral resolution.

3.1.4 Outlook

We have seen that the use of low-noise noise as masker does lead to different thresholds compared to Gaussian noise. This supports the idea that temporal fluctuations in Gaussian noise are a significant factor in auditory masking. Thus, low-noise noise may be an interesting stimulus also in the future to study the contribution of envelope fluctuations to masking.

For measurement techniques low-noise noise may also be of value because it is a signal that couples a low crest factor with a continuous spectrum. Although in hearing experiments, the bandwidth of low-noise noise is usually limited to at most that of one critical bandwidth to prevent that peripheral filtering reintroduces fluctuations in the envelope, for physical measurements this restriction may not exist and wideband low-noise noise may be used to put maximum wideband power in a system that somehow is restricted in its maximum amplitude.

3.2 Multiplied noise

While the generation of low-noise noise required intensive use of digital computers, the noise described in this section, multiplied noise, made its appearance as psychoacoustic stimulus already in the analog period [37]. It was a very convenient way to generate bandpass noises with tunable center frequency (see below) and these noises were therefore quite useful in spectral masking experiments in which noise maskers with variable center frequencies and steep spectral cutoffs were required [38,39]. The reason to discuss this stimulus in the context of this chapter is, however, based on its envelope properties, which allowed to test specific ideas about monaural and binaural hearing.

3.2.1 Definition

Multiplied noise, sometimes also called multiplication, or regular zero-crossing noise, is generated by directly multiplying a lowpass noise having a relatively low cutoff

frequency with a sinusoid of higher frequency. This operation is an example of modulation with suppressed carrier. The resulting signal is a bandpass noise with a center frequency equal to the frequency of the sinusoid. The bandwidth is given by twice the cut-off frequency of the lowpass noise. By tuning the frequency of the sinusoid, the center frequency of the multiplied noise is easily tunable, and such an operation does not require the use of digital computers.

3.2.2 Acoustic properties and perceptual insights

Multiplied noise has regular zero crossings with a rate equal to its center frequency. Its spectrum is symmetric around its center or carrier frequency: Above the carrier frequency, the spectrum is identical to the spectrum of the original lowpass noise, while the spectrum below the carrier is a mirrored version of this spectrum. Furthermore, the envelope distribution of multiplied noise is also different from that of Gaussian noise. While the envelope of Gaussian noise has a Rayleigh distribution, the distribution of the envelope of multiplied noise correponds to the positive half of a Gaussian distribution. This latter fact follows from the process of generation: the envelope of multiplied noise corresponds to the absolute value of the original lowpass noise time function. As a consequence, the envelope distribution of multiplied noise has its highest value at zero. In contrast, for Gaussian noise, the probability of zero envelope values is zero.

Due to the regular zero crossings, it is possible to add a sinusoid to multiplied noise with a fixed finestructure phase relation to the masker, if the sinusoidal frequency corresponds to the noise carrier frequency. If the sinusoid is added in phase, the resulting signal has still the same zero crossings as the noise alone, and overall, the envelope distribution is not much affected (see Fig. 11). When the sinusoidal signal

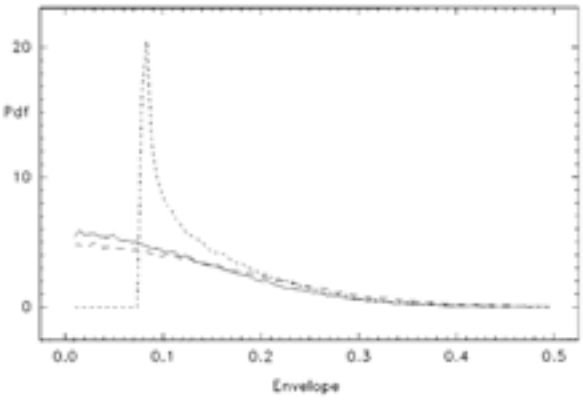


Figure 11. The envelope probability distribution is shown for a multiplied noise masker alone (solid line) and for a multiplied noise masker plus signal added in phase (long-dashed line), and with a fine-structure phase difference of $\pi/2$ (short-dashed line). The signal-to-masker ratio is 25 dB. Reused with permission from Ref. [40]. Copyright 1998, Acoustical Society of America.

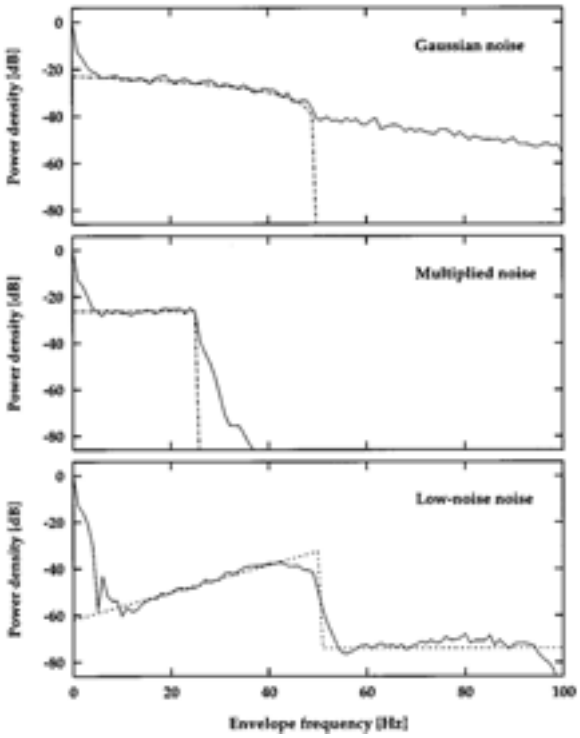


Figure 12. Idealized envelope spectra (dashed curves) and averaged spectra of Hilbert envelopes (solid curves) of Gaussian noise (upper panel), multiplied noise (middle panel), and low-noise noise (lower panel). Each averaged curve results from 1000 1-s-long realizations which were randomly selected from a 10-s-long noise buffer and then windowed with 50-ms Hanning ramps. Note that the power density is plotted on a logarithmic scale. Reused with permission from Ref. [41]. Copyright 1999, Acoustical Society of America.

has a phase difference of 90 degrees relative to the masker, the noise zero crossings appear when the sinusoidal waveform has its maxima (or minima). Therefore, in the resulting signal the low envelope values disappear. The dashed line in Fig. 11 shows this strong effect on the envelope distribution of a multiplied noise by adding a sinusoid with 90 degree phase difference.

Besides the distribution of the envelope values, also the envelope spectrum of multiplied noise differs from that of Gaussian noise. Due to its symmetric power spectrum, the envelope of ideal multiplied noise only has modulation frequencies up to half its bandwidth, while for Gaussian noise, intrinsic modulation frequencies range up to the whole bandwidth. Figure 12 shows envelope spectra, computed from 1-s long realizations, for 50-Hz wide narrowband noises with three different statistics: Gaussian noise (top panel), multiplied noise (middle panel) and low-noise noise (bottom panel). These different envelope spectra have been of great use in modelling amplitude modulation detection [41].

3.2.3 Role in hearing research and perceptual insights

Due to the many properties in which multiplied noise differs from Gaussian noise, it allows to investigate a great number of different psychoacoustic concepts. In the following, we want to give three examples:

- Role of masker fluctuations in off-frequency masking.
- Influence of the intrinsic envelope fluctuations of a noise on modulation detection.
- Effects of masker-signal phase relation in monaural and binaural on-frequency masking.

The spectral selectivity of the auditory system is often studied by using narrowband maskers [42]. A problem for the interpretation of masking data arises, when, due to the use of tonal masker, distortion products or temporal cues like beats influence the detection process [43]. In order to avoid the introduction of such "false cues", one often uses narrowband noise maskers instead [44]. It is implicitely assumed that this substituion has no other effect than eliminating the mentioned artifacts.

In a study by van der Heijden and Kohlrausch [45] this assumption underlying the power-spectrum model of masking [32,38] was tested for five different narrowband maskers with the same center frequency. The masker types were either a sinusoid, a Gaussian noise or a multiplied noise. The two noise signals had bandwidths of either 20 or 100 Hz. The five maskers thus differed in the amount of envelope variations: The sinuoid has a flat envelope, the two noises have envelopes with intrinsic fluctuations, with the multiplied noise having deeper envelope minima (see Sect. 3.2.2). The rate of fluctuation of the envelopes was another experimental parameter, because it is lower for the 20-Hz than for the 100-Hz masker.

In the experiments, these maskers had a center frequency of 1.3 kHz and were used to mask a sinusoidal test signal at 2 kHz. The masker levels were varied between 60 and 84 dB. Figure 13 shows the growth-of-masking functions for the five masker

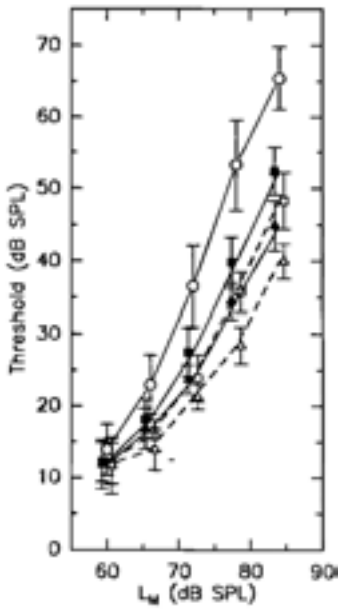


Figure 13. Simultaneous growth-of-masking functions measured with various narrow-band maskers centered at 1.3 kHz. The target was a 2-kHz tone. The five curves indicate the results for five different maskers: a sinusoid (open circles), a 100-Hz-wide Gaussian noise (filled squares), a 20-Hz-wide Gaussian noise (open squares), a 100-Hz-wide multiplied noise (filled triangles), and a 20-Hz-wide multiplied noise (open triangles). Data points measured with 20-Hz-wide maskers are connected with dashed lines. Averaged data of six subjects are plotted; error bars indicate ± one standard deviation across the values of the six subjects. Reused with permission from Ref. [45]. Copyright 1995, Acoustical Society of America.

types. Higher values of the threshold indicate a stronger effect of the masker on the audibility of the signal.

The data for the sinusoidal masker (circles, continuous line) indicate a highly nonlinear growth of masking. When the masker level is changed by 12 dB from 66 to 78 dB SPL, the threshold increases by 30.3 dB, corresponding to a slope of 2.5 dB/dB. The sinusoidal masker produces more masking than any of the other maskers. From all noise maskers, the least masking is obtained for the 20-Hz-wide multiplied noise masker (open triangles), while the strongest masking effect is seen for the 100-Hz-wide Gaussian masker. These data show thus the effect of two stimulus parameters on spectral masking: The more envelope minima occur in the masker, the lower the masked threshold (compare the three open symbols in Fig. 13). And, given a specific envelope distribution, faster fluctuations in the envelope, i. e., an increased bandwidth, lead to higher thresholds (compare open and filled symbols of the same type in Fig. 13).

Van der Heijden and Kohlrausch [45] emphasized in their conclusion the relevance of noise statistics for the interpretation of experimental data: "In summary, our data show that, particularly with high masker levels, attention should be paid to the exact nature of narrow-band stimuli used to mask a target at a frequency above the masker frequency [...] . These differences are by no means restricted to extremely slowly fluctuating maskers. On the contrary, masker fluctuations are relevant under many common experimental conditions, such as the 'low-frequency tail' of the psychoacoustic tuning curves. Unfortunately, in many published papers the influence of the type of noise used as masker (particularly multiplication noise versus Gaussian noise) has generally been neglected" ([45], p. 1806).

The envelope spectra of different noise types, shown in Fig. 12, suggested a critical test of the concept of modulation filter banks, which was developed in the Ph. D. thesis by Torsten Dau [36,46,47]. This concept provided a new view on modulation detection, and emphasized the role of the intrinsic envelope fluctuations of the carrier on the ability to detect sinusoidal amplitude modulations applied to this carrier. While for noise carriers, modulation detection is limited by intrinsic modulations of the carrier, for sinusoidal carriers, the limitation in detecting amplitude modulations comes from intrinsic properties of the auditory system [48]. An important element in this concept are modulation filters: a certain range of modulation frequencies falls into the same modulation filter, and modulation detection is possible if the applied modulation is sufficiently strong compared to the intrinsic modulations within the modulation filter centered on the test modulation.

In such a framework, one would expect very different modulation detection thresholds for noise carriers with different envelope spectra. For example, the modulation spectrum of low-noise noise has very little energy at low modulation frequencies and therefore, modulation detection should be much better at these modulation rates than, e. g., for Gaussian or also for multiplied noise. Dau et al. [41] tested this expectation by measuring modulation detection for the three noise types shown in Fig. 12.

In the left panel of Fig. 14, experimentally determined modulation thresholds are shown for the three noise types, all with a bandwidth of 50 Hz. First of all, we see that Gaussian and multiplied noises have the highest thresholds at low modulation

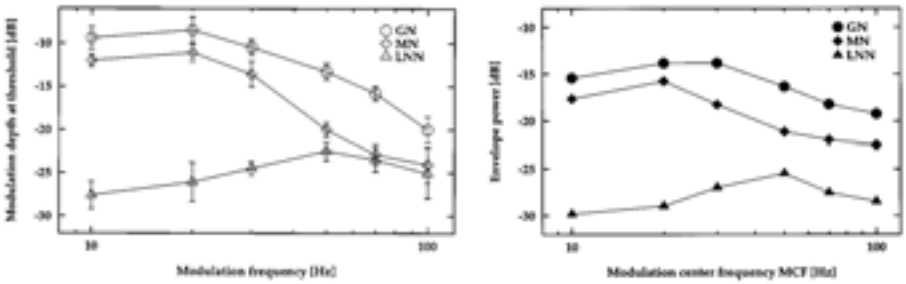


Figure 14. The two panels show experimental data for modulation detection (left) and the prediction of the modulation filterbank model developed in Ref. [36] (right). The symbols indicate the three different noise types. GN: Gaussian noise; MN: multiplied noise; LNN: low-noise noise. Reused with permission from Ref. [41]. Copyright 1999, Acoustical Society of America.

frequencies, while thresholds decrease for these two carrier types towards modulation frequencies corresponding to the carrier bandwidth of 50 Hz. In this range, thresholds for multiplied noise are clearly lower, which is in line with the analysis of the envelope spectra shown in Fig. 12. In contrast, the thresholds for low-noise noise are very low at the lowest modulation frequencies and increase towards 50 Hz. This increase directly resembles the increase in envelope power up to 50 Hz shown in Fig. 12.

The right panel shows the prediction of the modulation filterbank model that was originally developed for prediction of data obtained with Gaussian noise only [36,47]. The resemblance between data and model prediction is strong evidence that the basic ideas incorporated in the model capture properties of the hearing process in an appropriate way.

The last example for the potential of multiplied noise comes from binaural hearing. These experiments are based on the phase-locked interaction between masker and test signal and the possibility to create masking conditions with differences only in interaural time, or interaural level. These experiments will be discussed in the following section which focusses on binaural experiments.

4 Stimuli for research on binaural hearing

One of the main areas of psychoacoustic research at the DPI was binaural hearing, in which the particular consequences of differences between the acoustic waveforms at the two ears are studied. Binaural hearing addresses issues such as localization and lateralization of stimuli, and the wide area of binaural unmasking, the ability to detect a source in the presence of other background sources with the same or different spatial parameters. Binaural hearing requires that the signals from both ears are, somewhere in the hearing pathway, compared for coincidences and differences. Because such an interaction only takes place in neural centers far beyond the inner ear, it is very hard to directly test models of binaural hearing. Therefore, the interplay between models and the design of critical experiments is of particular relevance for advancing the understanding of human hearing in this area.

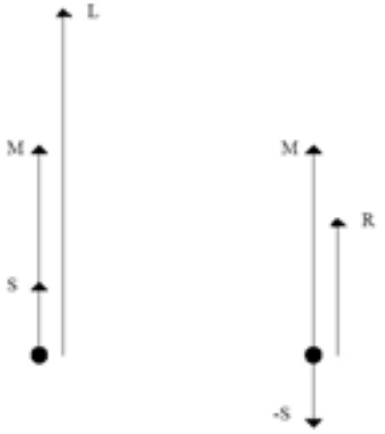


Figure 15. The vector representation of an N_0S_π stimulus. The masker M and signal S are added in the left ear with zero phase difference, resulting in the vector L, and subtracted in the right ear, resulting in the vector R. Only interaural intensity differences (IIDs) are present in the resulting binaural stimulus. Reused with permission from Ref. [40]. Copyright 1998, Acoustical Society of America.

Many of the headphone-based binaural experiments have used the paradigm of Binaural Masking Level Differences (BMLDs). BMLDs are indicators of a detection advantage in a specific binaural condition, relative to a reference condition without any interaural differences. Depending on signal parameters, this detection advantage can amount to more than 20 dB and it is commonly thought that this aspect of spatial hearing contributes to our ability to communicate in acoustically adverse environments (sometimes indicated with the term 'cocktail-party effect').

In this section, two signals types are described which both were introduced to evaluate specific model predictions about binaural unmasking.

4.1 Multiplied noise

As mentioned in the previous section, multiplied noise is a stochastic signal which has regular zero crossings. By adding a sinusoidal test signal with a frequency equal to the noise's center frequency, one has control over the relative phase between masker and signal, in a similar way as by adding two sinusoids with the same frequency. In the context of binaural experiments, this property can be extended to so-called N_0S_π conditions, in which the **N**oise masker has no interaural difference while the **S**ignal has an interaural phase difference of π. When a Gaussian noise is used as masker in such a condition, the addition of the sinusoidal signal introduces random fluctuations in both Interaural Time Differences (ITDs) and Interaural Intensity Differences (IIDs). With such a masker, it is not possible to asses the individual contributions of ITDs and IIDs to the process of binaural unmasking. In the following, we will describe how this is possible by using multiplied noise and sinusoidal signals with a fixed phase relation to the masker's waveform.

4.1.1 Acoustic properties and perceptual insights

In Fig. 15, the multiplied-noise masker M and a signal S are shown in vector representation. The two panels together show the right and the left signal for the condition N_0S_π. In the left panel, the angle between M and S is zero and, since the signal

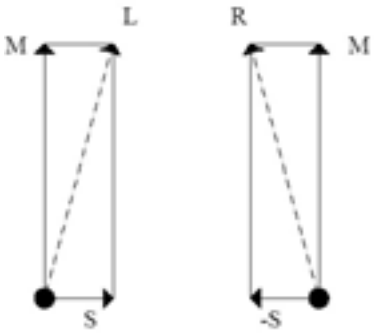


Figure 16. The vector representation of an N_0S_π stimulus. The masker M and signal S are added in the left ear with +90 degree phase difference, resulting in the vector L, and added in the right ear with -90 degree phase difference, resulting in the vector R. Only interaural time differences (ITDs) are present in the resulting binaural stimulus. Reused with permission from Ref. [40]. Copyright 1998, Acoustical Society of America.

is interaurally out of phase, in the right panel, the phase difference between masker and signal has to be π. The addition of the masker and signal in the left and the right panels results in two vectors, L and R, which together represent the binaural stimulus. The vectors L and R have in nearly all cases the same orientation. Only when the masker envelope M is smaller than the one of the signal S, L and R point in opposite directions. Basically, for this value of the masker-signal phase, the binaural stimulus contains only IIDs which vary in time at a rate proportional to the masker bandwidth.

In Fig. 16 a similar picture is shown as in Fig. 15, but now the masker M and the signal S are added with phase differences of +90 and -90 degrees. The resulting vectors L and R have the same length in all cases, only their orientation is different. This vector diagram describes a condition with no interaural intensity differences, and only time-varying interaural time differences.

These stimuli were used by van de Par and Kohlrausch [40] in a number of binaural experiments. We will here only discuss one of the measurements, which reveals nicely, how much ITDs and IIDs contribute to binaural unmasking at different frequencies. Figure 17 shows the experimental results for conditions with either only interaural intensity differences (circles) or only interaural time differences (squares). The difference between open and close symbols indicates the amount of binaural unmasking (BMLD). Two results become clear from this graph: when the binaural stimulus contains IIDs, the amount of binaural unmasking remains constant at all frequencies. When the stimulus contains interaural time differences, the binaural detection advantage is limited to low frequencies up to 1 kHz, and disappears completely at frequencies of 2 kHz and above. The details of the transition between 1 and 2 kHz seem to vary between the individual listeners. These data clearly demonstrate that BMLDs do not generally decrease at high frequencies, as is sometimes stated, but that the detection advantage is strongly linked to the cues that are available in the stimulus. This aspect will be further elaborated in the next section about transposed stimuli.

These results from binaural masking experiments agree very well with observations about the detection of static IIDs and ITDs using sinusoidal signals. The sensitivity

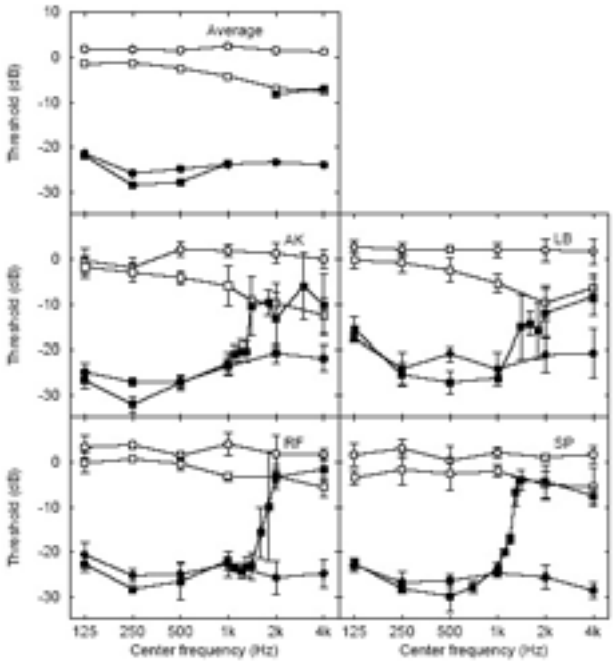


Figure 17. Thresholds for N_0S_0 (open symbols) and N_0S_π (closed symbols) are shown for masker-signal phases of 0 degree (cricles) and for phase values of 90 degrees (squares) as a function of center frequency. Thresholds were measured at a masker bandwidth of 25 Hz and are shown for four subjects in the lower four panels. Average results for the four subjects are shown in the upper panel. Reused with permission from Ref. [40]. Copyright 1998, Acoustical Society of America.

to changes in IID seems to be independent of center frequency and amounts to about 1 dB [49]. The sensitivity to ITDs contained in the finestructure of the signal decreases strongly above 1 kHz and is basically absent above 2 kHz [50,51].

4.2 Transposed stimuli

Transposed stimuli were introduced in the early 1990's to investigate the reasons why the BMLD for sinusoidal stimuli in broadband noise maskers decreased at high frequencies [52–54]. For the N_0S_π condition obtained with Gaussian noise, the BMLD is about 15 dB at 500 Hz, while it is reduced to 2 to 3 dB at 2 kHz. Two mechanisms were thought of to contribute to this reduction (see, e. g., Ref. [55]).

1. One contribution could come from the fact that auditory filters become wider at higher frequencies. Therefore, the rate of fluctuations of interaural differences at the output of an auditory filter will increase at higher frequencies. If one assumes that the auditory system is limited in its ability to follow those rapid changes, such an increased rate should have negative effects on the BMLD.

2. Another reason could lie in the loss of phase locking in the neural system at higher frequencies. This does imply that at frequencies above about 1.5 kHz, information about the finestructure of acoustic waveforms gets gradually lost. In consequence, at high frequencies the binaural system has only access to interaural differences in the envelope.

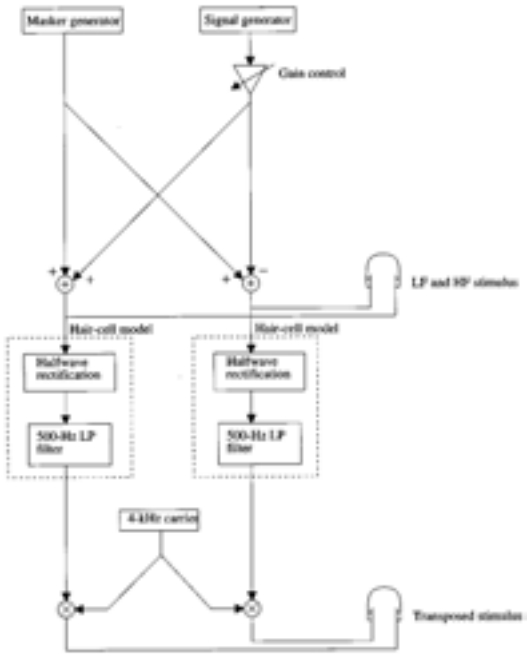


Figure 18. A scheme for the generation of transposed stimuli. The top part shows the generation of masker and signal waveforms, in this case for a binaural condition with the masker in phase and the signal interaurally out of phase. In the lower part, the low-frequency signals are halfwave rectified, lowpass filtered and multiplied with a high-frequency carrier. Reused with permission from Ref. [52]. Copyright 1997, Acoustical Society of America.

The question remained whether the differences in BMLDs at low and high frequencies were primarily a consequence of the loss of fine-structure information, or whether they reflected a structural difference of binaural processing between low and high frequencies. In order to distinguish between these explanations, transposed stimuli with a high carrier were introduced which contained in their envelope the same temporal information as a related low-frequency stimulus in its fine structure.

4.2.1 Definition

A transposed stimulus is derived from a narrowband low-frequency stimulus by the following operations. First, the low-frequency signal is halfwave rectified and lowpass filtered. In Ref. [52], a second-order lowpass with a cut-off frequency of 500 Hz was used. This signal is then used to modulate a high-frequency carrier, e. g., a sinusoid at 4 kHz.

In Fig. 18, these steps are shown schematically for stimuli as they are used in a masking experiment with adaptive adjustment of the signal level. In the top part of the figure, the generation and mixing of the low-frequency masker and the low-frequency signal are shown. The signal is added to the masker with a certain level, which is controlled by the adaptive procedure and adjusted by the gain stage. Depicted is the generation of an N_0S_π stimulus, for which the signal is inverted in one channel. After addition, the stimuli are ready for a standard low-frequency masking experiment. The bottom part shows the additional steps needed for the generation of a transposed stimulus.

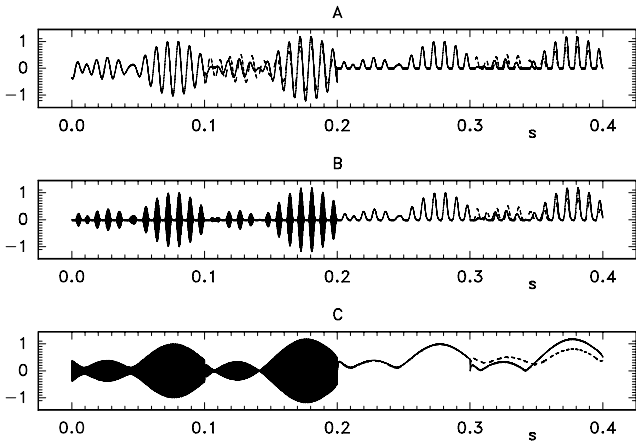


Figure 19. An example of three different N_0S_π stimuli before and after peripheral processing. Panel A shows a 125-Hz stimulus, panel B shows a transposed stimulus at 4 kHz, and panel C shows a standard 4-kHz stimulus. The intervals 0.0–0.1 s show the N_0 masker alone, the intervals 0.1–0.2 s show the N_0 masker plus the S_π signal at a signal-to-noise ratio of -10 dB, the intervals 0.2–0.3 s show the masker after peripheral processing, and the intervals 0.3–0.4 s show the combined masker and signal after peripheral processing. Reused with permission from Ref. [52]. Copyright 1997, Acoustical Society of America.

4.2.2 Acoustic properties

The construction of transposed stimuli followed the idea to build high-frequency signals that, after the initial stages of processing on the basilar membrane and the hair cells, have the same amount of temporal information as they are present in a standard low-frequency experiment. This property of transposed stimuli is depicted in Fig. 19. In each of the three panels, the left half of each panel shows acoustic waveforms as they are presented to subjects, and the right half, starting at 200 ms, shows these stimuli after the first stages of peripheral processing. The first 100 ms show the waveform of a noise masker alone, the section 100 to 200 ms contains the noise masker plus an S_π signal with a level 10 dB below the masker. The section 200 to 300 ms shows the masker after peripheral processing, and the last 100 ms show masker plus signal after peripheral processing. Panel A contains signals centered at 125 Hz, and panel B contains transposed signals with a center frequency of 4 kHz, derived from the low-frequency signal in panel A. Panel C finally shows standard high-frequency stimuli centered at 4 kHz. In this case, the masker is a narrowband noise centered at 4 kHz and the signal is a 4-kHz sinusoid.

By comparing panels A and B, we see that, indeed, after peripheral processing, the low-frequency channel in panel A and the high-frequency channel in panel B contain the same temporal information. In contrast, this information is strongly reduced in a standard high-frequency condition as shown in panel C.

Figure 20 shows the spectrum of a transposed stimulus, derived from a narrowband noise at 125 Hz. The spectral level is highest around the carrier frequency of 4 kHz.

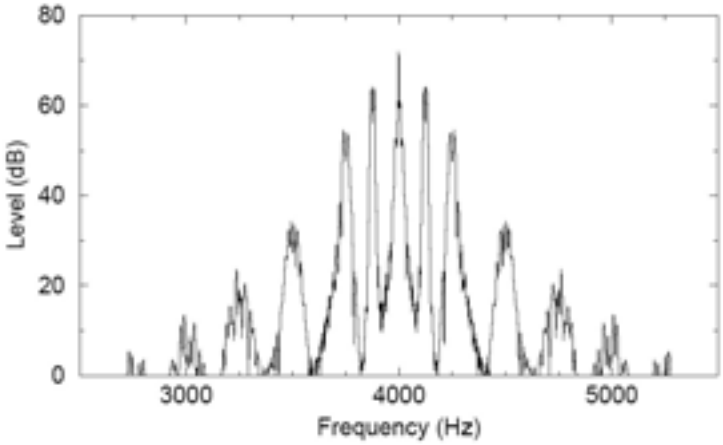


Figure 20. The spectrum of a 25-Hz-wide 125-Hz transposed stimulus. Most of the stimulus energy is around 4 kHz. The spectrum of the original low-frequency stimulus now occurs at 4125 Hz and the mirror image of that spectrum is at 3875 Hz. Additional peaks in the spectrum occur with spacings of 250 Hz. Reused with permission from Ref. [52]. Copyright 1997, Acoustical Society of America.

Spectral sidebands occur at spectral distances of 125, 250, 500, 750 etc. Hz. The spectrum of the original noise band is contained in the two sidebands directly to the left and the right of the carrier frequency.

4.2.3 Role in hearing research and perceptual insights

In Ref. [52], a number of experiments using transposed stimuli are described. Experiments were performed for transposed stimuli derived from 125-Hz and also from 250-Hz low-frequency conditions. Binaural conditions tested were N_0S_π and $N_\pi S_0$. For comparisons, thresholds were also obtained in the low-frequency condition, from which the transposed stimuli were derived, and for a standard high-frequency condition with a narrowband masker. Figure 21 shows experimental data for the condition N_0S_π for 125 Hz. Squares and circles indicate 125-Hz and 125-Hz transposed results, while diamonds indicate results for a standard 4-kHz measurement. The mean data show very clearly that 125-Hz and 125-Hz transposed stimuli give the same results, not only in the size of the BMLD, but also in the dependence on the masker bandwidth. In contrast, 4-kHz BMLDs in the standard condition are much lower.

This result is strong support for the idea that the major source of the decrease in BMLDs at high frequencies lies in the effect of peripheral processing on the available temporal information in the stimuli entering the binaural processing stage. If we manage to present very similar amounts of temporal information at a frequency with phase locking, e. g., 125 Hz and at a frequency without phase locking, e. g. 4 kHz, by using our modulation technique, then very similar amounts of binaural unmasking are measured. This observation thus supports an older hypothesis from Colburn and

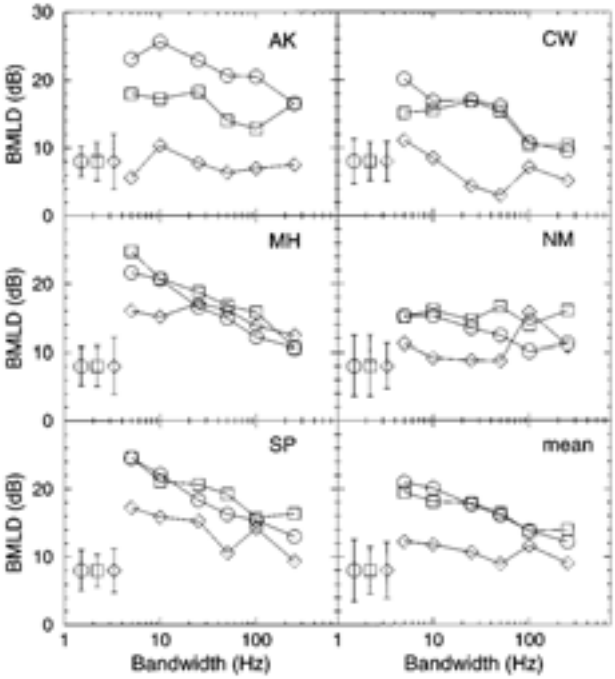


Figure 21. The N_0S_π BMLDs as a function of masker bandwidth for 125 Hz (circles), 125-Hz transposed (squares), and 4 kHz (diamonds). Five panels show data for individual subjects; the panel at the bottom right shows the average of the five subjects. In the bottom right panel, the three symbols with error bars indicate the averaged standard deviation of the mean BMLD for the individual subjects for the three conditions. In all other panels the three symbols with error bars indicate the averaged standard deviation of repeated measurements. Reused with permission from Ref. [52]. Copyright 1997, Acoustical Society of America.

Esquissaud [56] that the properties of the binaural processor are quite similar at low and high frequencies.

4.2.4 Later developments

The concept of transposed stimuli has been taken up in a wide range of hearing studies. Bernstein and Trahiotis [57,58] used transposed stimuli to measure the extent of laterality and the sensitivity to changes in interaural time differences, and compared their results with those obtained for standard low- and high-frequency stimuli. In line with the results found by van de Par and Kohlrausch [52], they observed great similarity in their measures between low-frequency and transposed stimuli. Oxenham et al. [59] used this type of stimuli in pitch perception experiments, in order to allow a separation of place and timing information, which is usually coupled when using sinusoidal signals.

Because the central assumption underlying the construction of transposed stimuli

is based on a "poor-man's" model of processing in the peripheral parts of the hearing system, the ultimate test for this concept had to come from physiological experiments. Griffin et al. [60] used transposed stimuli to measure ITD sensitivity in high-frequency neurons of the Inferior Colliculus of guinea pigs. Their study basically supported all assumptions and findings made in the original paper by van de Par and Kohlrausch [52] and these authors concluded: "Despite limitations in the temporal coding of neurons to high-frequency sounds, the results presented here demonstrate that under conditions in which binaural neurons receive appropriate spike patterns, sensitivity to ITDs conveyed by high-frequency stimuli can be equivalent to that observed in response to low-frequency stimuli. This suggests, as first conjectured by Colburn and Esquissaud (1976), that mechanisms underlying ITD sensitivity in low- and high-frequency channels of the auditory system are, to a first approximation, equivalent." ([60], p. 3477).

The latest "application" of transposed stimuli comes from the field of audiology. Long et al. [61] used a preprocessing based on the generation of transposed stimuli to test sensitivity to interaural differences in listeners wearing bilateral cochlear implants. They found that all their four subjects showed significantly improved performance when the signal was presented in a BMLD configuration, i. e. interaurally out of phase. This observation thus supports the potential of providing temporal information via the envelope of high-frequency carriers.

5 Conclusion

In this chapter, we have described a number of signal types with specific properties in their phase characteristics, their finestructure or their envelope, which have allowed detailed tests of ideas and concepts related to human hearing. In the text, we have concentrated on the acoustic properties and the consequential perceptual insights. But it also needs to be pointed out that the ability to generate these stimuli, and to develop advanced time-domain hearing models, depended heavily on the highly flexible and nearly bug-free software package that was created in the hearing groups at the DPI throughout the 1980's with contributions from many group members. SI, Signal processing Interactive, was the name for a program concept that allowed to work with digital signals like a pocket calculator works with numbers. In our opinion, the possibilities offered by this package were at least as essential for the scientific progress as were the growing insights in psychoacoustics and physiology.

The choice of signal types in this chapter is certainly biased towards those examples, to which we have ourselves contributed. The great potential of these stimulus types lies in the possibility to apply them easily in physiological experiments, and to use them as input to time-domain models which allow the processing of arbitrary signal waveforms [36,47,62–66]. This close interplay between psychoacoustics, physiology and modeling is one of the central themes of a conference series, the International Symposia on Hearing, which was initiated in 1969 by, among others, Manfred R. Schroeder and is since then organized every three years. This symposium never took place in Göttingen, but it can be seen as a late echo of the psychoacoustic research at the DPI that two recent editions of these symposia, the one in 2000 [67]

and the one in 2006 [68], were co-organized by scientists who received their initial academic training and were shaped in their scientific interests at the DPI.

Acknowledgements. The work described in this chapter reflects the close and fruitful cooperations that I (AK) had over the past more than 25 years with friends and colleagues at the DPI in Göttingen and later on in Eindhoven, where I was joined by SvdP in 1993. We both would like to thank all our colleagues for the creative atmosphere and great fun, which helped to generate some interesting scientific ideas. A part of this atmosphere was in 1991 exported to Eindhoven, and as can be seen from the reference list, this has become a similarly fruitful period. Finally, we would like to thank the senior staff at the DPI, in particular Udo Kaatze, for taking the initiative to publish this book.

References

[1] A. Kohlrausch, 'Erwin Meyers frühe Beiträge zur Psychoakustik', in *Fortschritte der Akustik, DAGA '07* (Deutsche Gesellschaft für Akustik, 2007), on CD ROM.
[2] E. Meyer, 'Über den derzeitigen Stand der Theorie des Hörens', Naturwissenschaften **12**, 358 (1947).
[3] H. Haas, 'Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache', Acustica **1**, 49 (1951).
[4] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, 'Comparative Study of European Concert Halls', J. Acoust. Soc. Am. **56**, 1195 (1974).
[5] S. Münkner, A. Kohlrausch, and D. Püschel, 'Influence of fine structure and envelope variability on gap-duration discrimination thresholds', J. Acoust. Soc. Am. **99**, 3126 (1996).
[6] B. R. Glasberg and B. C. J. Moore, 'Derivation of auditory filter shapes from notched-noise data', Hearing Research **47**, 103 (1990).
[7] H. Duifhuis, 'Audibility of High Harmonics in a Periodic Pulse', J. Acoust. Soc. Am. **48**, 888 (1970).
[8] M. R. Schroeder, 'Synthesis of Low-Peak-Factor Signals and Binary Sequences With Low Autocorrelation', IEEE Transact. Inf. Theor. **16**, 85 (1970).
[9] A. Kohlrausch and A. Sander, 'Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets', J. Acoust. Soc. Am. **97**, 1817 (1995).
[10] A. Kohlrausch and A. J. M. Houtsma, 'Edge pitch of harmonic complex tones', in *IPO Annual Progress Report* (1991), vol. 26, pp. 39–49.
[11] A. Kohlrausch and A. J. M. Houtsma, 'Pitch related to spectral edges of broadband signals', Phil. Trans. R. Soc. Lond. B **336**, 375 (1992).
[12] S. Mehrgardt and M. R. Schroeder, 'Monaural phase effects in masking with multi-component signals', in *Hearing – Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer Verlag, 1983), pp. 289–295.
[13] B. Smith, Master's thesis, Georg-August-Universität Göttingen (1984).
[14] H. W. Strube, 'A computationally efficient basilar-membrane model', Acustica **58**, 207 (1985).
[15] B. K. Smith, U. Sieben, A. Kohlrausch, and M. R. Schroeder, 'Phase effects in masking related to dispersion in the inner ear', J. Acoust. Soc. Am. **80**, 1631 (1986).
[16] A. Kohlrausch, 'Masking patterns of harmonic complex tone maskers and the role of the inner ear transfer function', in *Basic issues in hearing*, edited by H. Duifhuis, J. Horst, and H. Wit (Academic Press, Harcourt Brace Jovanovich, Publ., London, 1988), pp. 339–350.

[17] R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, 'The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold', J. Acoust. Soc. Am. **72**, 1788 (1982).
[18] P. I. M. Johannesma, 'The pre-response stimulus ensemble of neurons in the cochlear nucleus', in *Proceedings of the 2nd International Symposium on Hearing*, edited by B. L. Cardozo, E. de Boer, and R. Plomp (Institute for Perception Research, Eindhoven, 1972), pp. 58–69.
[19] R. D. Patterson, M. H. Allerhand, and C. Giguère, 'Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform', J. Acoust. Soc. Am. **98**, 1890 (1995).
[20] R. P. Carlyon and A. J. Datta, 'Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system', J. Acoust. Soc. Am. **101**, 3636 (1997).
[21] R. P. Carlyon and A. J. Datta, 'Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components', J. Acoust. Soc. Am. **101**, 3648 (1997).
[22] V. Summers and M. R. Leek, 'Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners', Hearing Research **118**, 139 (1998).
[23] V. Summers, 'Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes', J. Acoust. Soc. Am. **108**, 2307 (2000).
[24] A. J. Oxenham and T. Dau, 'Reconciling frequency selectivity and phase effects in masking', J. Acoust. Soc. Am. **110**, 1525 (2001).
[25] A. J. Oxenham and T. Dau, 'Towards a measure of auditory-filter phase response', J. Acoust. Soc. Am. **110**, 3169 (2001).
[26] A. Recio and W. S. Rhode, 'Basilar membrane responses to broadband stimuli', J. Acoust. Soc. Am. **108**, 2281 (2000).
[27] J. E. J. Hawkins and S. S. Stevens, 'The Masking of Pure Tones and of Speech by White Noise', J. Acoust. Soc. Am. **22**, 6 (1950).
[28] A. Langhans and A. Kohlrausch, 'Differences in auditory performance between monaural and diotic conditions. I: Masked thresholds in frozen noise', J. Acoust. Soc. Am. **91**, 3456 (1992).
[29] J. Pumplin, 'Low-noise noise', J. Acoust. Soc. Am. **78**, 100 (1985).
[30] W. M. Hartmann and J. Pumplin, 'Noise power fluctuation and the masking of sine signals', J. Acoust. Soc. Am. **83**, 2277 (1988).
[31] A. Kohlrausch, R. Fassel, M. van der Heijden, R. Kortekaas, S. van de Par, A. J. Oxenham, and D. Püschel, 'Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations', Acustica united with acta acustica **83**, 659 (1997).
[32] H. Fletcher, 'Auditory patterns', Rev. Mod. Phys. **12**, 47 (1940).
[33] B. C. Moore, J. I. Alcántara, and T. Dau, 'Masking patterns for sinusoidal and narrowband noise maskers', J. Acoust. Soc. Am. **104**, 1023 (1998).
[34] A. Kohlrausch and R. Fassel, 'Binaural masking level differences in nonsimultaneous masking', in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. Anderson (Lawrence Erlbaum Ass., 1997), chap. 9, pp. 169–190.
[35] S. van de Par and A. Kohlrausch, 'The role of intrinsic masker fluctuations on the spectral spread of masking', in *Forum Acusticum 2005, Budapest* (European Acoustics Association, 2005), pp. 1635–1640.
[36] T. Dau, B. Kollmeier, and A. Kohlrausch, 'Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers', J. Acoust. Soc. Am.

102, 2892 (1997).
[37] D. D. Greenwood, 'Auditory masking and the critical band', J. Acoust. Soc. Am. **33**, 484 (1961).
[38] R. D. Patterson, 'Auditory filter shape', J. Acoust. Soc. Am. **55**, 802 (1974).
[39] R. D. Patterson, 'Auditory filter shapes derived with noise stimuli', J. Acoust. Soc. Am. **59**, 640 (1976).
[40] S. van de Par and A. Kohlrausch, 'Diotic and dichotic detection using multiplied-noise maskers', J. Acoust. Soc. Am. **103**, 2100 (1998).
[41] T. Dau, J. Verhey, and A. Kohlrausch, 'Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers', J. Acoust. Soc. Am. **106**, 2752 (1999).
[42] T. H. Schafer, R. S. Gales, C. A. Shewmakers, and P. O. Thompson, 'The frequency selectivity of the ear as determined by masking experiments', J. Acoust. Soc. Am. **22**, 490 (1950).
[43] R. L. Wegel and C. E. Lane, 'The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear', Phys. Rev. **23**, 266 (1924).
[44] J. P. Egan and H. W. Hake, 'On the masking pattern of a simple auditory stimulus', J. Acoust. Soc. Am. **22**, 622 (1950).
[45] M. van der Heijden and A. Kohlrausch, 'The role of envelope fluctuations in spectral masking', J. Acoust. Soc. Am. **97**, 1800 (1995).
[46] T. Dau, *Modeling auditory processing of amplitude modulation*, Ph.D. thesis, Universität Oldenburg, Germany (1996).
[47] T. Dau, B. Kollmeier, and A. Kohlrausch, 'Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration', J. Acoust. Soc. Am. **102**, 2906 (1997).
[48] A. Kohlrausch, R. Fassel, and T. Dau, 'The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers', J. Acoust. Soc. Am. **108**, 723 (2000).
[49] A. Mills, 'Lateralization of High-Frequency Tones', J. Acoust. Soc. Am. **32**, 132 (1960).
[50] R. G. Klumpp and H. R. Eady, 'Some Measurements of Interaural Time Difference Thresholds', J. Acoust. Soc. Am. **28**, 859 (1956).
[51] J. Zwislocki and R. S. Feldman, 'Just Noticeable Differences in Dichotic Phase', J. Acoust. Soc. Am. **28**, 860 (1956).
[52] S. van de Par and A. Kohlrausch, 'A new approach to comparing binaural masking level differences at low and high frequencies', J. Acoust. Soc. Am. **101**, 1671 (1997).
[53] A. Kohlrausch, S. van de Par, and A. J. M. Houtsma, 'A new approach to study binaural interaction at high frequencies', in *Proceedings of the 10th international symposium on hearing, Irsee 94*, edited by G. Manley, G. Klump, G. Köppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore, New Jersey, London, Hong Kong, 1995), pp. 343–353.
[54] S. van de Par, *A comparison of binaural detection at low and high frequencies*, Ph.D. thesis, Technische Universiteit Eindhoven (1998).
[55] P. M. Zurek and N. I. Durlach, 'Masker-bandwidth dependence in homophasic and antiphasic tone detection', J. Acoust. Soc. Am. **81**, 459 (1987).
[56] H. S. Colburn and P. Esquissaud, 'An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli', J. Acoust. Soc. Am., Suppl. 1 **59** (1976).
[57] L. R. Bernstein and C. Trahiotis, 'Enhancing sensitivity to interaural delays at high frequencies by using 'transposed' stimuli', J. Acoust. Soc. Am. **112**, 1026 (2002).
[58] L. R. Bernstein and C. Trahiotis, 'Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed' stimuli', J. Acoust. Soc. Am. **113**, 3335 (2003).

[59] A. J. Oxenham, J. G. W. Bernstein, and H. Penagos, 'Correct tonotopic representation is necessary for complex pitch perception', Proc. Natl. Acad. Sci. **101**, 1421 (2004).
[60] S. J. Griffin, L. R. Bernstein, N. J. Ingham, and D. McAlpine, 'Neural Sensitivity to Interaural Envelope Delays in the Inferior Colliculus of the Guinea Pig', J. Neurophysiol. **93**, 3463 (2005).
[61] C. J. Long, R. P. Carlyon, R. Y. Litovsky, and D. H. Downs, 'Binaural unmasking with bilateral cochlear implants', J. Assoc. Res. Otolaryngol. **7**, 352 (2006).
[62] T. Dau, D. Püschel, and A. Kohlrausch, 'A quantitative model of the 'effective' signal processing in the auditory system: I. Model structure', J. Acoust. Soc. Am. **99**, 3615 (1996).
[63] T. Dau, D. Püschel, and A. Kohlrausch, 'A quantitative model of the 'effective' signal processing in the auditory system: II. Simulations and measurements', J. Acoust. Soc. Am. **99**, 3623 (1996).
[64] D. Breebaart, S. van de Par, and A. Kohlrausch, 'Binaural processing model based on contralateral inhibition. I. Model structure', J. Acoust. Soc. Am. **110**, 1074 (2001).
[65] D. Breebaart, S. van de Par, and A. Kohlrausch, 'Binaural processing model based on contralateral inhibition. II. Predictions as a function of spectral stimulus parameters', J. Acoust. Soc. Am. **110**, 1089 (2001).
[66] D. Breebaart, S. van de Par, and A. Kohlrausch, 'Binaural processing model based on contralateral inhibition. III. Predictions as a function of temporal stimulus parameters', J. Acoust. Soc. Am. **110**, 1105 (2001).
[67] J. Breebaart, A. Houtsma, A. Kohlrausch, V. Prijs, and R. Schoonhoven, *Physiological and psychophysical bases of auditory function* (Shaker Publishers, Maastricht, 2001).
[68] B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, and J. Verhey, eds., *Hearing – From sensory processing to perception*, Proceedings of the 14th International Symposium on Hearing (Springer Verlag, Berlin, Heidelberg, 2007).

Oscillations, Waves and Interactions, pp. 73–106
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-04-5

Sound absorption, sound amplification, and flow control in ducts with compliant walls

D. Ronneberger and M. Jüschke

Drittes Physikalisches Institut, University of Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. Efficient damping of narrow-band noise in ducts is commonly achieved with a lining which has a resonance tuned to the frequency band of the noise. However, with superimposed mean flow and with high quality factors of the resonators, large sound amplification (up to 30 dB) is observed even if only a short section (less than two diameters long) of the circular duct is provided with the lining. Jointly with the sound amplification, a considerable increase of the static pressure drop (by more than 100 % at high sound pressure amplitudes) along the lined duct section is observed. The most important experimental results will be reviewed and the physical mechanisms behind these phenomena are thoroughly discussed. The theoretical investigation of the wave propagation and of the stability of the flow within the resonator section happens to be an unexpectedly high challenge. The various common approaches based on mode decomposition and axial homogeneity of the flow result in dispersion relations which largely diverge from the experimental results. While a convective instability has been observed to cause the considered phenomena, the flow is predicted to be subject to absolute instability even if the interaction between the coherent and the turbulent instability waves are included by a first approximation. So we conjecture that the spatial development of the mean flow which for its part depends on the spatial development of the instability waves has to be taken into account.

1 Introduction

Sound propagation in channels plays an important role in many technical problems, e. g. when noise from machines or from the outside is propagated through air-conditioning duct systems or when the noise from turbofans is to be reduced. Duct acoustics is also an obligatory subject in lectures on technical acoustics, and the design of sound absorbing ducts is a problem, which is treated typically in engineering sciences. So, why – the reader might ask – are we concerned with this subject in a Physics Institute since many years? In fact, it is more than fifty years ago that Fridolin Mechel, a student of Erwin Meyer at that time, was instructed to investigate the performance of acoustically treated ducts that carry a mean flow. Only a few years earlier, Lighthill [1] had published his famous paper on the sound production in unsteady flows which has triggered extensive research work in aero-acoustics.

Mechel found out that the sound attenuation in a channel the wall of which is equipped with periodically spaced Helmholtz resonators is considerably reduced in the presence of mean flow and may even be turned into sound amplification at certain frequencies [2]. The relation between these frequencies, the flow velocity and the spacing of the resonators revealed a close analogy between the observed sound amplification and the amplification of electromagnetic waves in the so called travelling wave tube which exploits the interaction between the electromagnetic wave and an electron beam via a periodic structure [3]. However, the mechanism itself by which mean-flow energy is converted to sound energy remained unclear. Besides the sound amplification the excitation of loud tones was observed by Mechel when the resonators were undamped. The latter phenomenon is known also from isolated resonators the openings of which are exposed to grazing flow. A salient example of such self-excited pressure oscillations has occurred in a gas transport system where a pair of pipes with closed valves at the end branch off the main pipe in opposite directions. The pressure amplitudes were so large that the flow velocity had to be reduced in order to maintain safe operating conditions [4].

The phenomena observed by Mechel are two examples of aero-acoustic instabilities which are based on the so called Kelvin-Helmholtz instability of vortex sheets. Such vortex sheets form at the interface between the Helmholtz resonators and the interior of the flow duct. While the sound amplification and the self-excited tones are closely related to the inhomogeneity of the channel, namely to the spacing of the resonators and maybe to the width of their openings, similar phenomena are also caused by homogeneous compliant walls that bound internal and external flows.

The interest in the stability of fluid flow along compliant walls has been stimulated by the experiments which M. O. Kramer had performed to explain Gray's paradox: Gray [5], wondering about the fast swimming speed of some dolphin species, had calculated that the dolphin's muscles had to deliver a multiple of the mechanical power that is produced by the muscles of all other mammals unless the dolphins are able to control the flow in the boundary layer around their body to remain laminar. Kramer [6] speculated that the particular mechanical properties of the dolphin's skin stabilize the flow, and in fact, with special compliant coatings of his test bodies he obtained an appreciable drag force reduction. However trials to reproduce these spectacular results in other laboratories have failed. The seminal theoretical investigation by Benjamin [7] and an impressive number of subsequent investigations some of which are still in progress have nevertheless shown that the compliance of the wall has indeed a strong effect on the stability of the flow boundary layer (see, e. g., Refs. [8,9]). The flow is destabilized in most cases because of a whole zoo of instability modes that arise by the compliance of the wall, but under special circumstances a stabilization can be achieved, and there is little doubt meanwhile that the onset of turbulence can be delayed by appropriate coating of the wall.

In connection with sound propagation in acoustically treated flow ducts the existence of the mentioned kind of instability has first been noted in theoretical studies [10–16], however, there was no indication that these modes really exist in the turbulent environment of practical flow situations. Nevertheless, the excitation of instability modes at the leading edge of the compliant wall was taken into account, and it was assumed that the growing instability modes loose their coherence by nonlinear

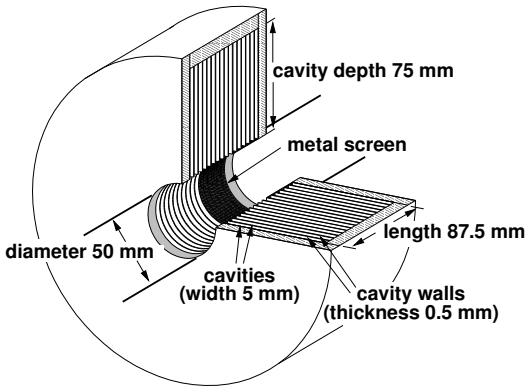


Figure 1. Dimensions of the mainly studied lining composed of 16 identical cavities: Resonance frequencies: First radial mode: 840 Hz, second radial mode: 3294 Hz, first azimuthal mode: 1113 Hz.

effects after some travelling distance and finally increase the turbulence of the flow. So the main issue of these theoretical studies was the scattering of incoming sound waves at the leading and possibly at the trailing edge of the compliant wall. Möhring and Eversman [12] and Quinn and Howe [14] controversially discussed the possibility that mean-flow energy can be converted to acoustic energy at the edges of lossless liners depending on the choice of the so called unsteady Kutta condition;while the real flow is approximated by potential flow, the Kutta condition is an assumption about the singularity of the potential at the edge (see Ref. [17]).

This discussion and a theoretical study by Koch and Möhring [13] who also highlight the influence of the Kutta condition on the amplitudes of the scattered sound waves was a reason to start an experimental investigation which should decide on the adequate choice of the Kutta condition. For this purpose a circular hard-walled flow duct was constructed to include a short section with a compliant wall. The lining was designed to have a well defined low-loss acoustic admittance. In addition, the response of the wall displacement to the pressure field should be as local as possible. Figure 1 shows a sequence of narrow annular cavities which meet these requirements and which are connected to the interior of the flow duct through fine-meshed metal screens to prevent the turbulent eddies to penetrate into the cavities.

The first experimental results however revealed that the original aim of the study was hardly reachable because the wave propagation in the lined duct section exhibits some significant differences to the assumptions which have been made in [12–14] and which had been anticipated to guide the evaluation of our experimental results [18,19]. So the study of the unsteady Kutta condition, which basically is a study of the dynamics of flow separation, was continued with other flow geometries and by other means [20–24], and as a kind of compensation, the flow through the lined duct section attracted our attention by a rather unexpected sound amplification that largely differed from the phenomenon observed and described by Mechel [2,3], and even more unexpectedly the static pressure drop along the lined duct section exhibited a strong

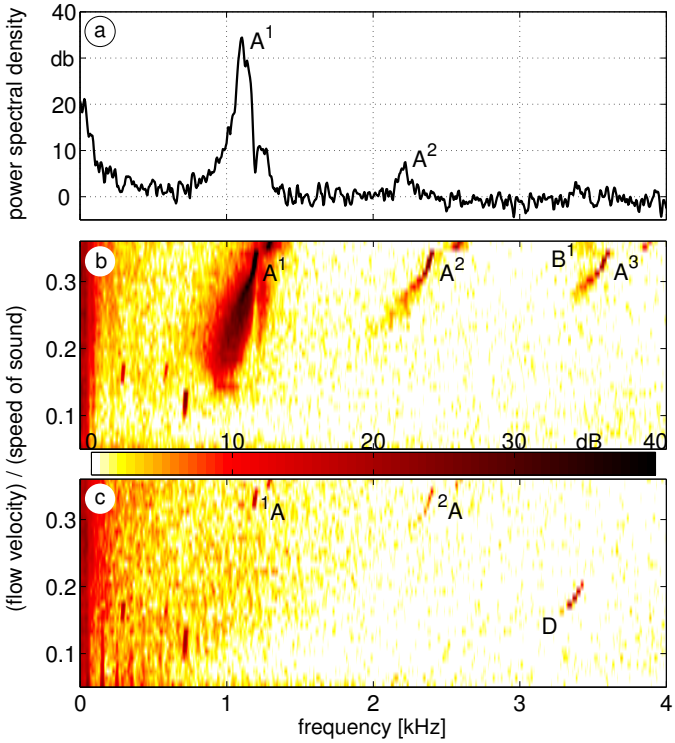


Figure 2. Power spectral density (psd) of the pressure downstream and upstream of the resonator section: (a) $\overline{U}/c = 0.27$, downstream; (b) difference to the median of the distribution of the psd, as a function of the flow velocity, downstream; (c) same as (b), but upstream.

dependence on the incident sound amplitude [25–28]. These phenomena have meanwhile turned out to be the effect of a strong seemingly convective instability of the turbulent flow in the lined duct section. Various experimental observations and some qualitative explanations of the phenomena will be presented in Sects. 2 and 3 while the attempts and the difficulties to understand the instability will be discussed in Sect. 4.

2 Experimental observations

2.1 Axisymmetric mode

2.1.1 Turbulent pressure fluctuations, sound transmission and static pressure drop

The sound amplification and the acoustic influence on the static pressure occur at frequencies slightly above the first radial resonance frequency of the cavities. In contrast to the original study with frequencies well below the resonance frequency, most of the experiments to be reviewed in the following have therefore been performed

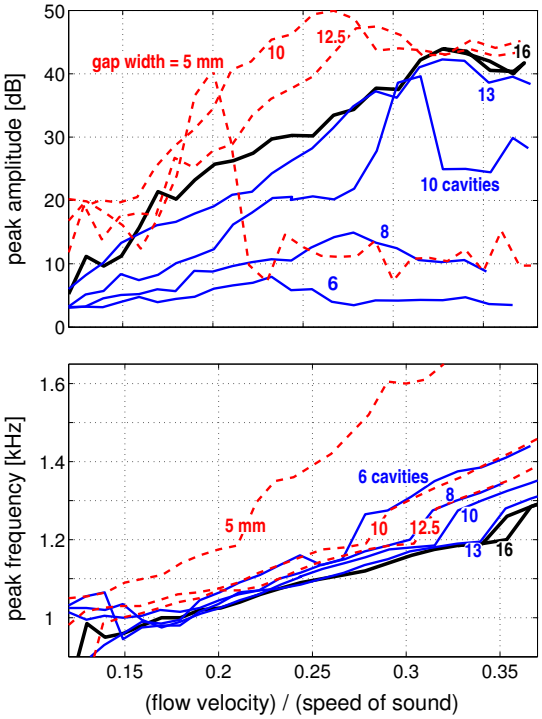


Figure 3. Amplitude and mid-frequency of the A^1-peak of Fig. 2(b) as functions of the flow velocity. The number of cavities has been varied while the full cross-section of the pipe was open to the flow (blue and black solid curves, denoted by the number of cavities), and the cross-section of the duct was reduced by insertion of a coaxial cylindrical body while the number of cavities (16) remained constant (red dashed curves, denoted by the gap width between the central body and the duct wall).

with a low resonance frequency (840 Hz) corresponding to the dimensions of the cavities that are shown in Fig. 1. The lowest cut-on frequency of higher-order modes in the rigid pipes upstream and downstream of the resonator section is 4 kHz. So for frequencies in the order of the resonance frequency all higher-order modes are evanescent and only the fundamental axisymmetric mode can propagate.

We first became aware of the sound amplification by some faint narrow-band whispering superimposed on the flow noise in the laboratory. The pressure spectra (Fig. 2) measured far upstream and far downstream of the resonator section reveal that this sound originates in the resonator section and is radiated mainly in the direction of the flow. The sample spectrum in Fig. 2(a) shows a few prominent peaks on top of an otherwise rather constant power spectral density, and the development of these peaks with increasing flow velocity[†] $\overline{U}$ is depicted in Figs. 2(b) and (c) where the

[†]Although the compressibility of the air is not essential for the considered phenomena, the average of the flow velocity over the cross-sectional area of the pipe $\overline{U}$ is normalized to

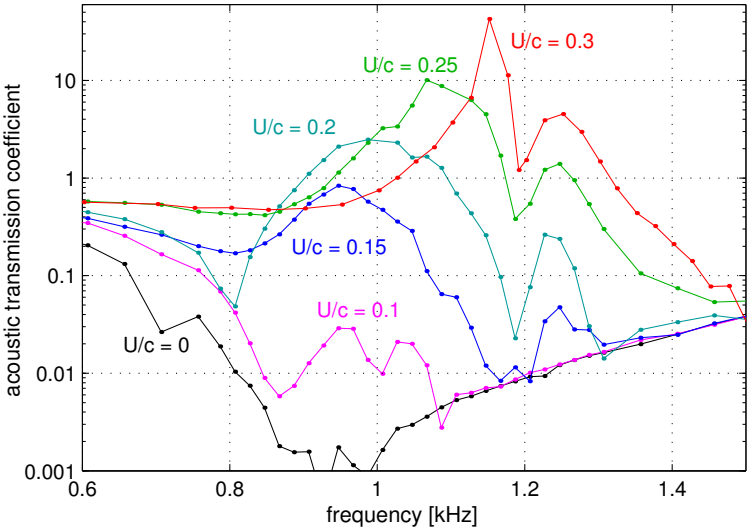


Figure 4. Acoustic transmission coefficient of the unmodified resonator section as functions of the freqency for various flow velocities. The incident sound wave travels in the direction of flow. Resonance frequency: 840 Hz.

median of the distribution of the power spectral density (psd) is taken as 0 dB. The enhancement of the psd at low frequencies is not caused by the resonator section but is due to the internal noise of the flow facility.

The most prominent peak in Fig. 2(a) is marked by A^1. The amplitude of this peak becomes very large at high flow velocities so that higher harmonics (A^2 and A^3) appear in the spectra. At these high flow velocities the pressure fluctuations are also radiated to the opposite direction of flow (^{1}A and ^{2}A in Fig. 2(c)). The peak B^1 in Fig. 2(b) slightly above the frequency of the second radial resonance of the cavities (3.3 kHz) is possibly based on the same mechanism as the peak A^1, while the peak D in Fig. 2(c) has been identified by Brandes [29,30] to be the result of another type of instability that will be reconsidered in Sect. 4.1.5.

The amplitude and the mid-frequency of the A^1-peak are plotted as functions of the flow velocity in Fig. 3 for different geometric parameters of the resonator section. While the dimensions of the cavity remain unmodified, the number of cavities is varied, and besides the circular duct cross-section also various annular cross-sections have been investigated characterized by the width of the annular gap between a central cylindrical body and the pipe wall. The peak frequencies continuously increase when the flow velocity is increased except for an interval between 1200 Hz and 1260 Hz which is always skipped. The jump of the frequency occurs at $\overline{U}_{\mathrm{jump}}/c \approx 0.35$ for the

the sound speed if definite values are given, however the term 'Mach number' is avoided; in fact we have not succeeded in finding an adequate and practical reference speed which remains constant during a typical experiment and does not depend, e. g., on the frequency or the acoustical admittance of the lining.

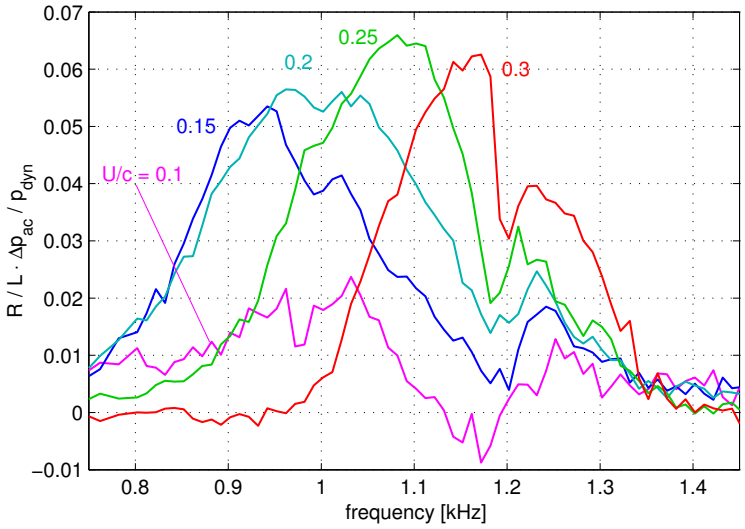


Figure 5. Sound-induced static pressure drop Δp_{ac} along the unmodified resonator section as a function of the frequency with various flow velocities. The average pressure gradient $\Delta p_{\mathrm{ac}}/L$ within the resonator section is normalized with the dynamic pressure per radius p_{dyn}/R.

unmodified resonator section depicted in Fig. 1, and a hysteresis is associated with the jump of the frequency (black curve in Fig. 3). The maximum peak amplitudes are reached slightly below $\overline{U}_{\mathrm{jump}}$ which strongly depends on the geometric parameters of the resonator section.

Besides the pressure spectra the sound transmission through the lined duct section has been investigated. A multi-microphone method with two independent sound fields has been used [31] for this purpose. Figure 4 shows the transmission coefficient for the sound propagating in the mean flow direction. At frequencies close to the resonance frequency and with $\overline{U} = 0$ the sound transmission is very effectively blocked, and this blockage becomes even more effective by mean flow in the opposite direction of the sound propagation (not shown in Fig. 4). However, with mean flow in the direction of sound propagation, the transmission coefficient considerably increases and becomes even much greater than unity, at high flow velocities. Considering the dependency on the freqency and on the flow velocity, obvious agreement is found between the sound amplification and the spectral peak (A^1) downstream of the resonator section. Therefore we conclude that the A^1-peak is caused by the amplification of the axisymmetric component of the turbulent pressure fluctuations that enter the lined duct section.

In conjunction with the sound amplification Krause [25] and Brandes [26,29] have observed an increase of the static pressure drop along the lined duct section, i. e. an increase of the mean wall shear stress in the resonator section. The dependency of the sound-induced pressure drop on the frequency is plotted in Fig. 5 for various

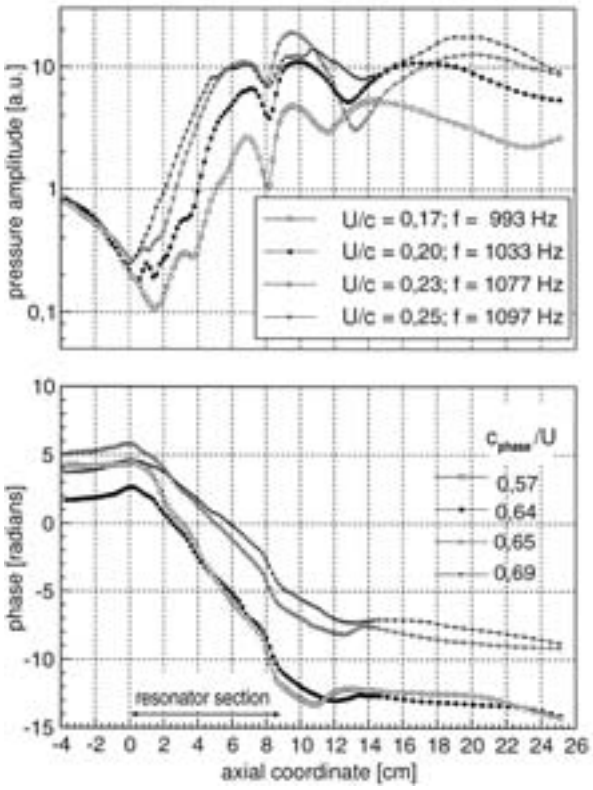


Figure 6. Amplitude and phase of the pressure oscillation close to the wall of the unmodified resonator section as functions of the axial coordinate; the pressure field was excited by an incident sound wave. The flow velocity U was varied and the respective frequency f was adjusted to the maximum of the transmission coefficient for each flow velocity. From Ref. [29].

flow velocities, and the sound was introduced into the first cavity by means of three loudspeakers, in this experiment. The greatest effect is obtained slightly above the frequency of maximum sound amplification. In fact, the pressure drop and the sound transmission exhibit a very similar dependency on the frequency and the flow velocity. These phenomena have recently been observed also by Aurégan et al. [32] with cavities which are much narrower than in our experiment; the ceramic skeleton of a catalytic converter has been used by these authors for that purpose.

The resonator section can be used as a valve which is contolled by the amplitude of superimposed sound and the response time of which (a few milliseconds) is given by the width of the frequency interval of sound amplification. Lange and Ronneberger [27] have demonstrated this possibility by active suppression of the sound transmission through the resonator section at frequencies up to ca. 200 Hz. For this purpose the pressure drop was adjusted to exactly follow the sound pressure in front

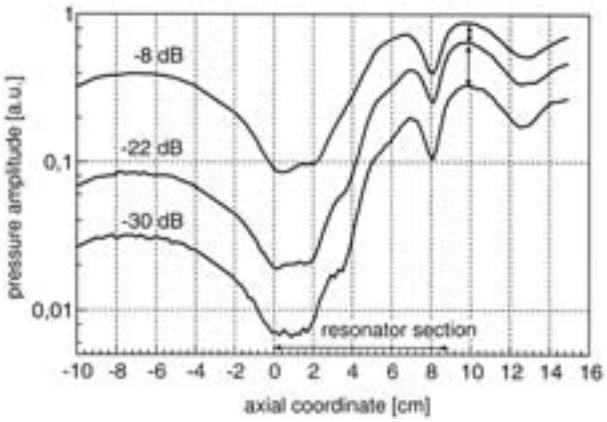


Figure 7. Pressure amplitude close to the wall of the unmodified resonator section with different amplitudes of the incident sound wave. $\overline{U}/c = 0.2$, f=1033 Hz. From Ref. [29].

of the resonator section. The relatively involved relation between the far upstream measured signal of the broad-band noise that was to be cancelled and the necessary amplitude of the controlling sound signal was determined and adjusted by means of an adaptive filter.

2.1.2 Pressure amplitude and static pressure in the resonator section

Considering the various types of instability in flow ducts with compliant walls, as mentioned in the introduction, we have come to the conclusion that the sound amplification is caused by such an instability. We assume that a convective instability mode is excited by the incident sound at the leading edge of the resonator section. The amplitude grows while the instability wave propagates through the lined duct section, and from the large pressure oscillation at the end of the lined duct section a large-amplitude sound wave is radiated into the rigid duct which is connected to the rear of the resonator section. The spatial distribution of the pressure oscillation in the resonator section supports this hypothesis. First of all it turns out that the pressure field is composed of various modes which differ by the wavenumber and by the radial dependence of the pressure. So rather involved interference patterns may occur, e. g. the axial gradient of the phase may be positive along the axis of the duct and be negative along the wall. The axial distributions depicted in the Figures 6 – 8 can be considered to be comparatively clear. Summarizing we arrive at the following conclusions [29,33]:

(i) Besides evanescent modes which are excited at the ends of the resonator section the pressure field contains a wave which exhibits all properties of an instability mode.

(ii) The phase velocity of the instability wave is $\mathcal{O}(\overline{U}/2)$.

(iii) The parameters that determine the propagation of the instability mode strongly vary along the axial coordinate; in particular the rate of exponential growth of the wave decreases with the travelling distance and may even become negative (Fig. 8

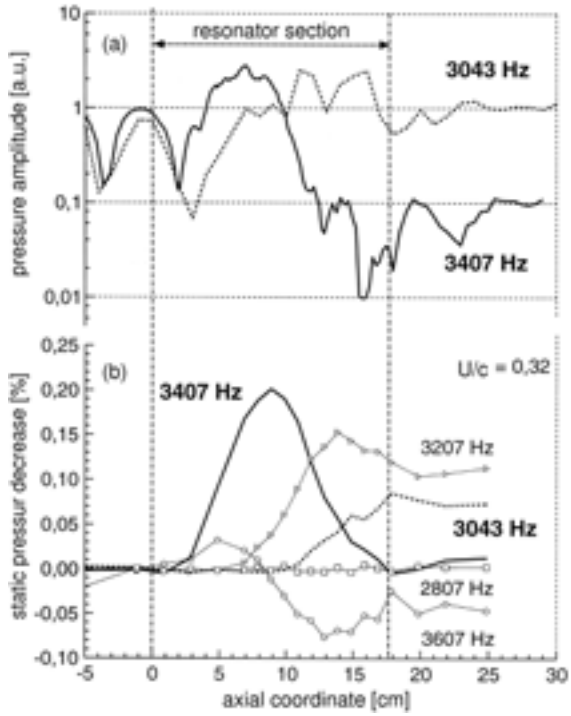


Figure 8. Pressure amplitude (a) and sound-induced static pressure decrease (b) within a resonator section consisting of 32 cavities with the first radial resonance at 2.94 kHz. The static pressure decrease is normalized to the pressure drop without sound irradiation (24 mbar). From Ref. [29].

shows an example). In such cases the acoustic transmission coefficient decreases when the length of the lined duct section is further increased, and since the spatial development of the instability wave depends on the frequency, the maximum of the transmission coefficient is reached at different frequencies for different lengths of the lined duct section (the frequency decreased when the length is increased).
(iv) The amplitude of the instability wave saturates at high amplitudes of the incident sound wave (see Fig. 7). So also the transmission coefficient decreases if a certain sound amplitude is exceeded.

Besides the oscillating part of the pressure also the sound-induced decrease of the static pressure was measured in several cases, and a close relationship between both the components of the pressures has been found:
(i) While the pressure amplitude increases as a function of the axial coordinate, the static pressure decreases, and vice versa (Fig. 8).
(ii) It may even happen at high sound frequencies that the static pressure increases and that the flow resistance of the resonator section decreases with sound irradiation.
(iii) While the sound amplitude increases, the pressure drop becomes noticeable only when the transmission coefficient starts to decrease.

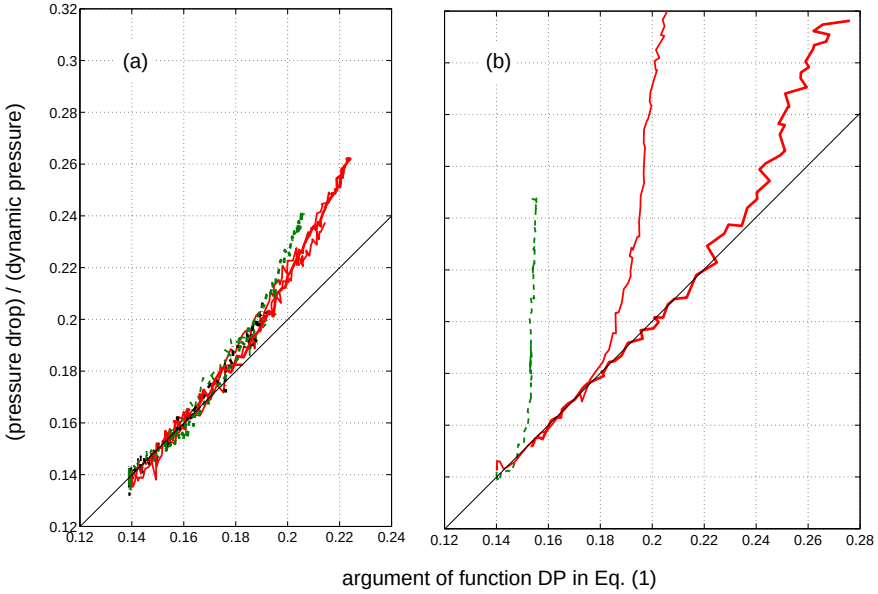


Figure 9. Relation between the pressure drop and the pressure amplitude at the rear end of the resonator section according to Eq. (1). (a) $m = 0$; p_{ac} = transmitted sound pressure amplitude; $\overline{U}/c = 0.2$ (thin curves), frequencies: 907 Hz (black, dash-dotted), 1007 Hz (red, solid), 1087 Hz (green, dashed); $\overline{U}/c = 0.25$ (thick curves), frequencies: 1007 Hz (black, dash-dotted), 1087 Hz (red, solid), 1127 Hz (green, dashed). (b) $m = 1$, p_{ac} = pressure amplitude in the backmost cavity; $\overline{U}/c = 0.2$ (thin curves), frequencies: 1207 Hz (red, solid), 1307 Hz (green, dashed); $\overline{U}/c = 0.25$ (thick curve), frequency: 1207 Hz (red, solid).

(iv) The pressure drop Δp along the lined duct section and the amplitude p_{ac} of the transmitted sound wave are related by

$$\frac{\Delta p}{p_{\mathrm{dyn}}} = \mathcal{DP}\left\{\left[\frac{p_{\mathrm{tb}}}{p_{\mathrm{dyn}}}(\overline{U})\right]^2 + \left[\frac{\kappa(f,\overline{U})\cdot p_{\mathrm{ac}}}{p_{\mathrm{dyn}}}\right]^2\right\}, \tag{1}$$

wherein p_{dyn} is the dynamic pressure. The quantities $p_{\mathrm{tb}}(\overline{U})$ and $\kappa(f,\overline{U})$ are fitted to the experimental data except for a free common factor which is adapted such that $\mathcal{DP}\{\cdots\}$ becomes the identity function for small sound pressure amplitudes; $\kappa(f,U)$ increases with increasing frequency and with decreasing flow velocity. As seen from Fig. 9(a), $\mathcal{DP}\{\cdots\}$ does not differ very much from the identity function also at large sound amplitudes. However, only a few experimental data were suited for this evaluation, so the universality of $\mathcal{DP}\{\cdots\}$ may be questioned. In fact, when higher-order mode sound irradiation (see next Section) is applied to the resonator section, strongly nonlinear and all but universal relations are found between Δp and p_{ac}^2. Fig. 9(b) shows several examples where p_{ac} is the pressure amplitude in the backmost cavity of the resonator section.

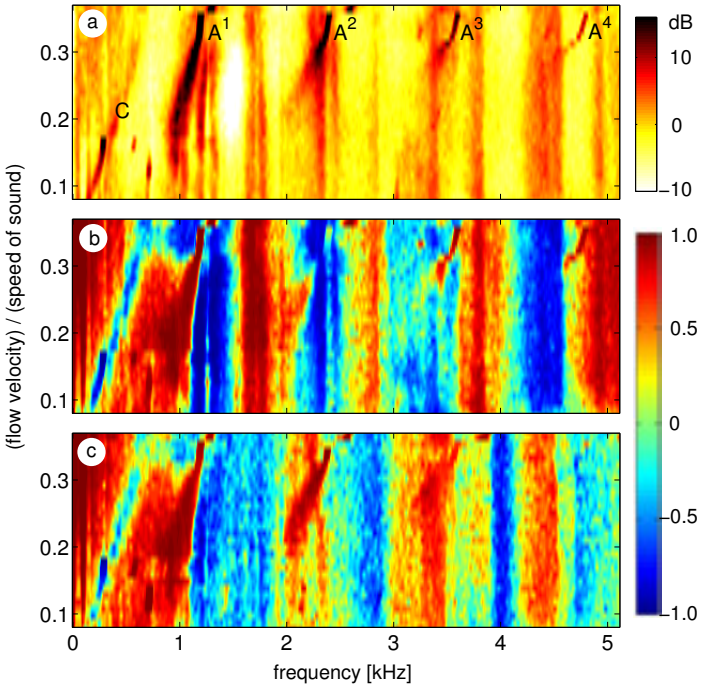


Figure 10. Power spectral density and coherence of the pressure in the backmost cavity of the unmodified resonator section plotted as a function of the frequency and of the flow velocity. (a) power spectral density (normalized to the background spectrum); (b) coherence between two locations the azimuthal coordinates of which differ by $\Delta\varphi = 180^{\circ}$; (c) like (b) with $\Delta\varphi = 120^{\circ}$.

2.2 Higher-order modes

Regarding the possible technical application of the sound-induced static pressure drop, the excitation and the effect of higher-order modes have been studied. It is anticipated that some of these propagate in the resonator section, but are evanescent in the rigid ducts so that the mostly unwelcome propagation of the amplified controlling sound is suppressed. We are particularly interested in circumferential modes which are easier to handle than radial modes. Then the amplitudes of the flow velocity and the pressure depend on the azimuthal coordinate φ according to $(\hat{u}, \hat{v}, \hat{p}) \sim \exp(im\varphi)$ wherein m is the azimuthal wavenumber and $|m|$ is the azimuthal order of the mode.

Because of the evanescence of the considered modes we have to abandon the radiated pressure fluctuations and the acoustic transmission coefficient as sources of information in this study; instead we analyze the pressure in the backmost cavity. The spectum of the pressure is more complex, and it contains more peaks than the radiated pressure spectra shown in Fig. 2; in addition the 'background spectrum' on top of which the peaks are observed exhibits a stronger dependency on the frequency than the radiated pressure. The background spectrum was therefore approximated

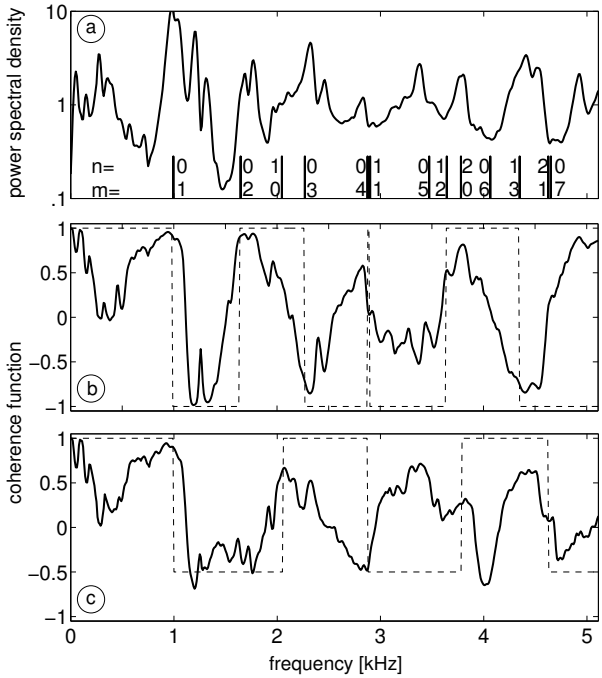


Figure 11. Average of the data of Fig. 10 over $0.13 \leq \overline{U}/c \leq 0.25$. The cut-on frequencies of the various modes in the resonator section are marked on the bottom of Fig. (a) where the radial order of the modes are denoted by n, and the expected coherence functions (see text) are shown as dashed curves in Figs. (b) and (c).

by a polynomial which was fitted to the logarithm of the power spectral density (psd), and only the difference between the actual psd and this 'background spectrum' is presented as a function of the frequency and the flow velocity in figure 10a. As in Fig. 2(b) the peaks A^1, A^2, and A^3 are the the most outstanding ones also in Fig. 10(a) where the fourth harmonic A^4 is included.

In addition quite a few less prominent peaks appear in Fig. 10(a) which have no correspondence in Fig. 2(b). These peaks are therefore supposed to originate from higher-order modes which are evanescent in the rigid pipes. With one exception (C), the mid-frequencies of the additional peaks do not depend on the flow velocity. So the clearness of these peaks can be increased by averaging the spectra over a range $\overline{U}/c = 0.13 \cdots 0.25$ of flow velocities, within which the less prominent peaks more or less dominate over the A-peaks. The average is shown in Fig. 11(a), and moreover the cut-on frequencies of the higher-order modes in the resonator section are marked on the frequency axis; the cut-on frequencies are only weakly dependent on the flow velocity and therefore have been computed for air at rest. A steep increase of the power spectral density is found at the low-order cut-on frequencies. So it is conjectured that the contribution to the pressure fluctuations by a definite mode is particularly large at frequencies just above the respective cut-on frequency.

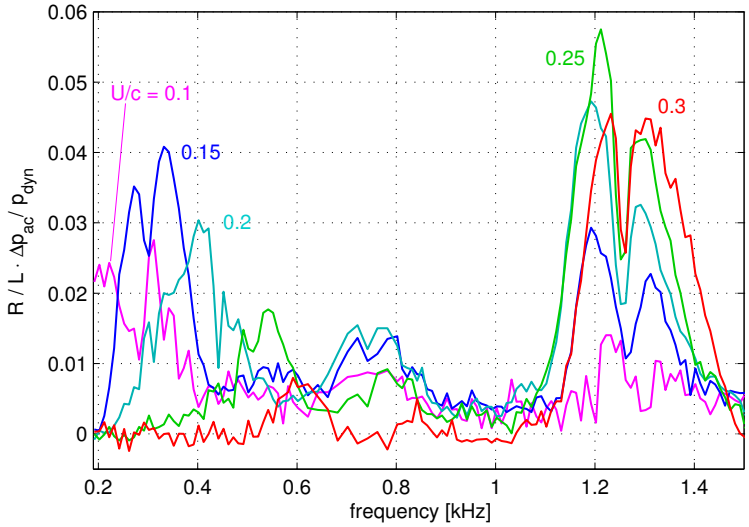


Figure 12. Increase of the pressure drop along the lined duct section effected by acoustical excitation of the $m = 1$ mode, as a function of the frequency and for various flow velocities.

This hypothesis is supported by the power cross-spectral density between locations the azimuthal coordinates of which differ by $\Delta\varphi = 180^{\circ}$ and by $\Delta\varphi = 120^{\circ}$, respectively. First of all, the imaginary parts of the cross-spectra are expected to vanish, since otherwise one of the two directions of rotation of the constituent modes (m positive or negative) would be preferred. Nevertheless, in a few cases significant imaginary parts of the power cross-spectral density are encountered, the origin of which was not pursued, however. Instead the imaginary parts are simply ignored in the following. If the modes $\hat{p}_m \exp(im\varphi)$ contained in the pressure field $\hat{p}(\varphi)$ are incoherent (what is to be expected) we obtain

$$\Re\langle\hat{p}(\varphi_1)\hat{p}^*(\varphi_2)\rangle = \langle|\hat{p}_0|^2\rangle + \sum_{m=1}^{\infty}(\langle|\hat{p}_m|^2\rangle + \langle|\hat{p}_{-m}|^2\rangle)\cos[m(\varphi_1 - \varphi_2)] \qquad (2)$$

wherein $\hat{p}(\varphi)$ is the Fourier transform of one possible pressure signal measured at the azimuthal location φ, and $\langle\cdots\rangle$ is the average over all possible realizations. So all the modes with identical order $|m|$ contribute to the real part of the cross-spectrum in proportion to their mean square amplitudes.

The real part of the coherence function, i. e. the real part of the ratio between the power cross-spectral and (auto-)spectral densities is plotted as a function of the flow velocity in Figs. 10(b) and (c) for $\Delta\varphi = 180^{\circ}$ and $\Delta\varphi = 120^{\circ}$, respectively. Also the average of the coherence function over the flow velocity ($\overline{U}/c = 0.13 \cdots 0.25$) was computed and was depicted in Figs. 11(b) and (c) together with coherence functions computed according to Eq. (2) with the assumption that the pressure field consists of one single mode namely the one which belongs to the nearest lower cut-on frequency, in each case. Obviously this assumption is not too bad, at least for low-order modes.

The conspicuous low-frequency peaks (C) in Fig. 10(a) the mid-frequencies of which are proportional to the flow velocity and which are absent in Fig. 2(b), are obviously caused by an antisymmetric pressure distribution according to Figs. 10(b) and (c). With the assumption that this pressure distribution is due to a single mode $|m|$ we infer from the measured coherence functions that $\cos(m{\cdot}180^{\circ}) < 0$ and $\cos(m{\cdot}120^{\circ}) < 0$, according to Eq. (2), so that $|m| = 1$, or $|m| = 5, 7, \cdots$. In Figs. 11(b) and (c) this mode manifests itself by the broad dips at low frequencies. There are some indications (Sect. 4.1.2) that the considered mode is a hydrodynamic mode which cannot exist but with non-zero mean flow.

The effect of the higher-order modes on the pressure drop was investigated only with $|m| = 1$. In order to excite this mode, three loudspeakers with an azimuthal spacing of 120° and driven at 120° phase difference were connected to the first cavity of the resonator section. Also the axisymmetric mode $m = 0$ can be excited by this means, namely by driving the loudspeakers with identical phases. First of all it turns out that these acoustically excited modes, particularly $m = 1$, reach the backmost cavity only within certain frequency bands. For frequencies up to ca. 2 kHz, these bands more or less coincide with the frequency bands within which the coherence function ($\Delta\varphi = 180^{\circ}$, Fig. 10(b)) is negative, respectively positive when the $m = 0$ mode is propagated through the resonator section. There is however some overlap between the respective pass-bands, thus, e. g., both the modes propagate within the band denoted by C in Fig. 10.

As expected, the static pressure can be noticeably influenced by the $m = 1$ mode only within the pass-bands of this mode. This is shown in Fig. 12 for various flow velocities. With low flow velocities the low-frequency hydrodynamic mode C is more effective than the mode which propagates above the first acoustical cut-on frequency of the $|m| = 1$ modes, however for $\overline{U}/c > 0.2$ the pressure drop is higher with this 'acoustical' $|m| = 1$ mode. Interestingly, this latter mode is effective in practically the same freqency range as the $m = 0$ mode the effects of which have been denoted by A and have been described in the previous Section 2.1. A comparison between the Figures 5 and 12 shows that nearly the same acoustically induced pressure drop is achieved with both these modes. A closer inspection reveals that some kind of interference seems to occur between the modes around 1.2 kHz: the pressure drop is high with the $m = 1$ mode when it is low with the $m = 0$ mode and vice versa.

3 Physical mechanisms

3.1 Interaction between the mean and the fluctuating parts of the flow

3.1.1 Momentum transport by flow oscillations and stability of the mean flow

In order to study the physical mechanisms behind the sound amplification and the acoustical control of the static pressure, we describe the interaction between the mean and the fluctuating parts of the flow in the common way: the flow velocity $\boldsymbol{u} = (u, v, w)^t$, the pressure p, and the density of mass ρ are decomposed into mean and fluctuating parts, $\boldsymbol{u} = \overline{\boldsymbol{u}} + \boldsymbol{u}'$, $p = \overline{p} + p'$, $\rho = \overline{\rho} + \rho'$, and are substituted in the Navier Stokes equation (conservation of momentum). Then the same type of average,

e. g. temporal, is taken of the equation, as was applied in the decomposition of the field quantities. As a result the mean transport of momentum which is described by the tensor $\overline{\rho \boldsymbol{u}\boldsymbol{u}^t}$ is composed of two parts: $\overline{\rho \boldsymbol{u}\boldsymbol{u}^t} = \rho\ \overline{\boldsymbol{u}'\boldsymbol{u}'^t} + \rho\ \overline{\boldsymbol{u}}\ \overline{\boldsymbol{u}}^t$ wherein the fluctuation of the density is disregarded. So the fluctuation of the velocity contributes to the mean transport of momentum by the covariance of the velocity components. The term $-\rho\ \overline{\boldsymbol{u}'\boldsymbol{u}'^t}$ is also known as Reynolds stress tensor because it contributes to the balance of forces between the volume elements in the same way as the pressure and the viscous friction. Particularly the Reynolds shear stress $\overline{\tau}'' := -\rho\overline{u'v'}$ which exceeds the viscous shear stress $\overline{\tau}_\mu := \mu\ \partial\overline{u}/\partial y$ by orders of magnitude in turbulent flow, plays the dominant role in the development of the mean flow profile.

On the other hand, the mean velocity profile is crucial to the development of the Reynolds tensor. To understand this we consider the hypothetical case that the fluctuating part of the flow consists of a single mode of a small-amplitude oscillation so that the Navier Stokes equation can be linearized. Additionally we confine our consideration to the simple incompressible flow $\overline{\boldsymbol{u}} = [U(y), 0, 0]^t$ in a 2D channel where we can use the wave ansatz

$$\{u', v', p'\}(x, y, t) = \{\hat{u}, \hat{v}, \hat{p}\}(y) \cdot \exp[i(\alpha x - \omega t)] \tag{3}$$

(with angular frequency ω, wavenumber α, axial and wall-normal space coordinates x and y). Then the Orr-Sommerfeld equation is obtained for the amplitude $\hat{v}(y)$ of the wall-normal component of the flow velocity. This is a fourth-order differential equation which reduces to the second order Rayleigh equation if the effects of viscosity are disregarded:

$$\frac{\mathrm{d}^2\hat{v}}{\mathrm{d}y^2} = \left(\alpha^2 + \frac{\mathrm{d}^2U/\mathrm{d}y^2}{U - \omega/\alpha}\right)\cdot\hat{v}\,; \qquad \hat{u} = \frac{i}{\alpha}\frac{\mathrm{d}\hat{v}}{\mathrm{d}y}\,; \qquad \hat{p} = \frac{i}{\alpha}\left[\rho\frac{\mathrm{d}U}{\mathrm{d}y}\hat{v} - i(\omega - \alpha U)\rho\hat{u}\right] \tag{4}$$

and for the sake of completeness the Rayleigh equation has been supplemented by relations for $\hat{u}$ and $\hat{p}$. Together with the boundary conditions at the walls an eigenvalue problem is established which generally has a number of solutions $\omega = \Omega_n(\alpha)$. Each of these solutions is the dispersion relation of a mode which may contribute to the fluctuating part of the flow. It is obvious from the Rayleigh equation (4) that the dispersion relations and consequently $(\hat{u}, \hat{v})_n$ as well as the Reynolds shear stress $\overline{\tau}''_n = -\rho\frac{1}{2}\Re\{\hat{u}\cdot\hat{v}^*\}_n$ depend on the mean flow profile, in particular on $\mathrm{d}^2U/\mathrm{d}y^2$. Furthermore it should be noted that $\overline{\tau}''_n$ crucially depends on the phase difference between $\hat{u}$ and $\hat{v}$ and hence can appear also as a negative-friction force.

3.1.2 Exchange of energy between mean flow and flow oscillation

The Reynolds shear stress is essential for the exchange of energy between the mean flow and the flow oscillation. The rate at which energy per volume is transferred from the mean to the oscillating part of the flow is given by the product of shear stress and shear rate $\overline{\tau}''\cdot\mathrm{d}U/\mathrm{d}y$ (angular momentum × angular velocity). From the incompressible Navier Stokes equations, one obtains

$$\overline{\tau}''\cdot\frac{\mathrm{d}U}{\mathrm{d}y} = -\rho\overline{u'v'}\cdot\frac{\mathrm{d}U}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}x}U\frac{\rho}{2}\overline{u'^2 + v'^2} + \frac{\mathrm{d}}{\mathrm{d}y}\overline{p'v'} + \epsilon_\mu\,. \tag{5}$$

Herein ϵ_μ comprises the transport of energy due to viscosity, i. e., diffusion and dissipation, which will be disregarded, to begin with. Thus the power extracted from the mean flow (left-hand side of Eq. (5)) is fed into the streamwise flux of kinetic energy and into the wall-normal energy transport due to the pressure-velocity correlation (first and second term on the right-hand side, respectively). While the wall-normal energy transport balances the dissipation of energy in the walls and therefore is irreversible, the kinetic energy contained in oscillation of the flow can be returned to the mean flow if the Reynolds shear stress becomes negative. Under this condition the first term on the right-hand side is negative as well, meaning that the amplitude of the oscillation decreases in the mean-flow direction, and we will see from Eq. (7) that the static pressure may increase under the same condition.

3.1.3 Static pressure gradient

The supply or the recovery of mean flow energy in a homogeneous channel cannot change the kinetic energy of the mean flow when the velocity profile is assumed to remain constant along the axial coordinate x. Hence the exchange of energy must be balanced by the work done by the pressure gradient on the flow. For the sake of simplicity, the following formulae are restricted to the 2D channel (half-width R) and to the channel with circular cross-section (radius R). The wall shear stress $\overline{\tau}_\mathrm{w}$ is constant over the circumference $\partial(S)$ of the channel in both these cases. So the balance between the forces due to the pressure gradient, due to the wall friction, and due to the acceleration of the fluid yields

$$-\frac{\mathrm{d}\overline{p}}{\mathrm{d}x} = \overline{\tau}_\mathrm{w}\frac{\partial(S)}{S} + \frac{\mathrm{d}}{\mathrm{d}x}\int_S \rho U^2 \frac{\mathrm{d}S}{S}\,. \tag{6}$$

The axial pressure gradient $\mathrm{d}\overline{p}/\mathrm{d}x$ is assumed to be constant over the cross-section which is a good approximation except for extreme axial alteration of the mean velocity profile $U(r,x)$. We are particularly interested in the first term on the right-hand side of Eq. (6) which is due to the wall shear stress and which will be denoted by $\mathrm{d}\overline{p}_\tau/\mathrm{d}x$. After some analysis the work done by $\mathrm{d}\overline{p}_\tau/\mathrm{d}x$ on the flow, namely $\int_S(\mathrm{d}\overline{p}_\tau/\mathrm{d}x)U\,\mathrm{d}S$, can be equated with

$$-\int_S \frac{\mathrm{d}\overline{p}_\tau}{\mathrm{d}x}U\mathrm{d}S = \overline{\tau}_\mathrm{w}\overline{U}\partial(S) \quad \underbrace{\overset{\overline{\tau}_\mu \ll \overline{\tau}''}{=} \int_0^R \left[\overline{\tau}\frac{\mathrm{d}U}{\mathrm{d}y}\right]\frac{S}{R}\mathrm{d}y}_{\text{2D channel}} \quad \underbrace{\overset{\overline{\tau}_\mu \ll \overline{\tau}''}{=} \int_0^R \left[\overline{\tau}\left(-\frac{\mathrm{d}U}{\mathrm{d}r}\right)\right] 2\pi r\mathrm{d}r}_{\text{circular duct}} \tag{7}$$

wherein

$$\overline{\tau}(r) = \overline{\tau}''(r) + \overline{\tau}_\mu(r) = \overline{\tau}_\mathrm{w}\frac{r}{R} \tag{8}$$

has been used which applies to the circular duct as well as to the 2D channel with $r = R - y$. So, according to the middle part of Eq. (8), the right-hand side of Eq. (7) comprises the power that is fed into the oscillation of the flow according to Eq. (5) and the power that is dissipated by the mean viscous shear stress $\overline{\tau}_\mu$. Furthermore it turns out that the wall shear stress is a measure of the considered energy transfer.

3.2 Qualitative explanation of the observed phenomena

With these basic physical mechanisms that have been summarized in the previous sections we can speculate about the physics behind the phenomena that have been described in Sect. 2. To begin with, the development of the turbulent flow in the resonator section will be considered in the presence of the hypothetical convective instability. Obviously part of the turbulent fluctuations of the incoming flow is amplified by the instability according to the spectra shown in Figures 2 and 10. The radial distribution of the Reynolds shear stress which is necessarily involved in the amplification of the fluctuations (Eq. (5)) is expected to differ from the distribution in the fully developed channel flow (Eq. (8)). This difference inevitably leads to an alteration of the velocity profile with the consequence that the stability of the flow and particularly the dispersion relation of the considered instability wave depends on the axial coordinate. Besides the development of the mean flow profile also the considerable nonlinear interaction between the growing fluctuations contributes to the change of the dispersion relation. This interaction will be described by means of the oscillating component of the Reynolds shear stress in Sect. 4.2.1.

The axial growth rate of acoustically excited small-amplitude instability waves has indeed been found to depend considerably on the travelling distance (Sect. 2.1.2). This indicates that a significant readjustment of the velocity profile already occurs within the short lined duct section – its length L has been varied between $1.8R \leq L \leq 7.2R$ in the experiments. Apart from the flow instability it is expected that the permeability of the wall causes a sudden increase of the wall shear stress when the flow enters the lined duct section (to be discussed in more detail in Sect. 4.2.1). This is confirmed by the pressure drop along the lined duct section which is closely related to the wall shear stress according to Eq. (6): also with short resonator sections the pressure drop increases by nearly one order of magnitude (compared to impermeable smooth wall). So it can be assumed that the readjustment of the mean flow profile starts at the wall and immediately at the leading edge of the lined duct section.

With regard to the acoustically induced increase of the pressure drop we consider the effect of the instability wave that is excited by the irradiated sound. Both, the kinetic energy of the fluctuating part of the flow and the Reynolds shear stress will increase in proportion to the square of the amplitude of the instability wave since the turbulent fluctuation and the acoustically excited oscillation are uncorrelated. With low sound amplitudes the acoustical part of the Reynolds shear stress is small compared to the turbulent part. So the mean flow remains unaffected, and the propagation of the acoustically excited instability wave is exclusively governed by the developing flow which has been described in the previous paragraph. Only when the acoustical and the turbulent parts of the Reynolds stress become comparable, the flow will be affected and both, the static pressure drop and the acoustic transmission coefficient will depend on the sound amplitude. In fact, the thresholds at which the dependency on the sound amplitude becomes noticeable, are fairly identical for the two effects.

The pressure drop can be calculated by integration of Eq. (6) over the length L of the resonator section, provided the second term on the right-hand side of the equation, i. e. the effect of alteration of the velocity profile can be disregarded. This

is indeed justified because the pressure drop is determined by the static pressure far ahead and far behind the resonator section where the velocity profile is fully adapted to the smooth rigid wall of the pipe. The pressure drop which would arise along a smooth rigid pipe between the measuring points was subtracted from the measured pressure difference. So only the pressure drop due to the additional wall shear stress in the resonator section is recorded. With Eq. (7) wherein $\overline{\tau}\,\mathrm{d}U/\mathrm{d}r$ has been replaced by means of the Eqs. (5) and (8), one obtains for the circular pipe

$$\begin{aligned}\Delta p &= -\int_0^L \frac{\mathrm{d}\overline{p}_\tau}{\mathrm{d}x}\,\mathrm{d}x = -\frac{1}{\overline{\overline{U}}}\int_0^L\int_0^R\left(\overline{\tau}''\frac{\mathrm{d}U}{\mathrm{d}r} + \overline{\tau}_\mu\frac{\mathrm{d}U}{\mathrm{d}r}\right)\frac{2r\mathrm{d}r}{R^2}\,\mathrm{d}x \\ &= \left[\int_S \frac{U}{\overline{\overline{U}}}\left(\frac{\rho}{2}\overline{u'^2+v'^2}\right)\frac{\mathrm{d}S}{S}\right]_0^L + \\ &\quad + \int_0^L \frac{\partial(S)}{S}\,\frac{\overline{p'v'}_\mathrm{w}}{\overline{\overline{U}}}\,\mathrm{d}x + \frac{1}{\overline{\overline{U}}}\int_0^L\int_S\left(\epsilon_\mu - \tau_\mu\frac{\mathrm{d}U}{\mathrm{d}r}\right)\frac{\mathrm{d}S}{S}\,\mathrm{d}x\,. \quad (9)\end{aligned}$$

First we discuss the most simple case and omit all dissipative effects which are comprised in the second and the third term on the right-hand side of Eq. (9). Then Δp is given by the axial change of the flux of kinetic energy contained in the fluctuation of the velocity. The acoustically excited part of the kinetic energy density $(\rho/2)\overline{u'^2_\mathrm{ac}+v'^2_\mathrm{ac}}$ is proportional to the square of the amplitude of the instability wave, and since the acoustic energy density is very small at the leading edge of the resonator section, the acoustically induced pressure drop Δp_ac is proportional to the square of the pressure amplitude at the rear end the of resonator section which on its part determines the pressure amplitudes of the transmitted sound and in the backmost resonator cavity. So the experimental results shown in Fig. 9 and described by Eq. (1) can be understood for medium sound amplitudes.

The dissipative terms in Eq. (9) exhibit the same quadratic dependency on the amplitude of the instability wave as the first term on the righthand side of the equation as long as the dispersion relation of the instability wave is not too much affected by the nonlinearity of the wave propagation. However, when the growth of the instability wave becomes saturated at high sound amplitudes (see Fig. 7), the kinetic energy at the rear end of the resonator section is more affected by the nonlinearity than the dissipative terms in Eq. (9). So while the pressure drop further increases with the incident sound amplitude the radiated pressure and the pressure in the backmost cavity begin to stagnate, i. e. the data points more and more lie above the 45° line in Fig. 9. In addition one has to take into account that the ratio between the amplitudes of the pressure and the velocity is influenced by the considered nonlinearity as well, however there is no simple way to predict how the function $\mathcal{DP}\{\cdots\}$ in Eq. (1) is affected by this implication of the nonlinearity.

4 Theoretical approaches to the wave propagation in the lined duct

In order to substantiate the physical explanation of the considered phenomena we need a deeper insight into the properties of the instability wave. Various aspects of the wave propagation in the lined duct have been studied for this purpose [34,35].

As usual the oscillating part of the flow field is decomposed into independent modes each of which is characterized by the cross-sectional distribution of the oscillating quantities $\{\hat{u}, \hat{v}, \hat{p}\}_n$ and by the dispersion relation $\omega = \Omega_n(\alpha)$; one such example has already been discussed in connection with the Rayleigh equation (4). In the case of uniform mean flow ('flat' velocity profile) the whole set of the modes is a complete base for the description of the unsteady part of the field in the duct. A further reason why a flat velocity profile is assumed in many theoretical studies is the advantage that both, the mean and the oscillatory part of the flow have a potential in this case, and that the distributions of the velocity and the pressure $\{\hat{u}, \hat{v}, \hat{p}\}_n$ can be described by well-known analytic functions in simple cross-sectional geometries like the 2D and the circular channel. The eigenvalue problem by which $\Omega_n(\alpha)$ and $\{\hat{u}, \hat{v}, \hat{p}\}_n$ are determined is then reduced to the solution of a transcendental equation.

4.1 Disregarding of shear stress and uniform mean flow

4.1.1 Boundary condition at the wall

The flat velocity profile exhibits a discontinuity at the wall, i. e., the boundary layer is replaced by an infinitesimal vortex layer. So it is not trivial which of the unsteady quantities remain continuous when passing through the vortex layer and how, consequently, the boundary condition, i. e. the acoustical admittance of the wall, is to be translated to the field in the interior of the channel. In any case the pressure has to be continuous since otherwise the lateral pressure gradient and the lateral acceleration of the fluid elements would tend to infinity. In addition the lateral deflection of the fluid elements on both sides of the vortex layer are assumed to be equal. This assumption however is justified only in the absence of shear stress as will be shown in Sect. 4.2.1.

The unsteady deflection of the wall and of the vortex layer, which sticks to the wall, so to speak, can be clearly defined. However the extension of this term to the flow field is problematic even if an infinitesimal oscillatory flow is superimposed on the mean flow. The vortex layer can be described as a set of streamlines which are determined by the integration of the field of flow directions. Thus the unsteady deflection of a streamline generally depends on the arbitrary starting point of the integration. To avoid this problem some local information about the temporal and the spatial development of the oscillating streamline is assumed to exist such that the relation between the transverse component $\boldsymbol{v}'$ of the flow velocity and the hypothetical deflection $\boldsymbol{\eta}'$ of the streamline can be used to define the deflection

$$\boldsymbol{v}'(\boldsymbol{x}, t) = \frac{\partial \boldsymbol{\eta}'(\boldsymbol{x}, t)}{\partial t} + U(\boldsymbol{x}) \frac{\partial \boldsymbol{\eta}'(\boldsymbol{x}, t)}{\partial x} =: \frac{\mathrm{D}\boldsymbol{\eta}'(\boldsymbol{x}, t)}{\mathrm{D}t} . \tag{10}$$

Herein $\boldsymbol{x} = (x, y, z)^t$ is the 3D position vector, and the x-direction is aligned with the mean-flow direction ($\overline{\boldsymbol{u}}(\boldsymbol{x}) = [U(y, z), 0, 0]^t$). Then $\boldsymbol{v}'$ and $\boldsymbol{\eta}'$ are 2D vectors in the (y, z)-plane. The abbreviation $\mathrm{D}/\mathrm{D}t := \partial/\partial t + U\partial/\partial x$ stands for the temporal derivative in a frame of reference that is moving at velocity U. Thus, from Eq. (10) and from the assumption that the deflection $\boldsymbol{\eta}'$ of the streamlines is solenoidal, i. e. that the compressibility of the fluid can be ignored on the infinitesimal scale of the

vortex layer, a jump condition for the wall-normal velocity component can be derived. In a wave-like field according to Eq. (3)

$$\hat{v}_\perp = \frac{\breve{\omega}}{\omega}\hat{v}_\mathrm{w} \tag{11}$$

is obtained wherein $\hat{v}_\perp$ and $\hat{v}_\mathrm{w}$ are the wall-normal velocity components of the wall and the fluid elements, respectively, and $\breve{\omega} = \omega - \alpha U$ is the Doppler-shifted frequency in the frame of reference of the fluid elements. So while the boundary condition at the wall is a local relation between the wall-normal velocity and the pressure for zero mean flow, it becomes non-local due to the dependency on α in the presence of mean flow.

4.1.2 Dispersion relations

In view of the sound amplification in the resonator section Großer [34,36] has studied the potential of various common approaches to the theoretical treatment of the wave propagation in lined ducts. Unfortunately none of these approaches results in dispersion relations that at least approximately exhibit the observed dependencies on the frequency and the flow velocity. One such approach will be discussed in the following in order to identify the immanent problems which still have to be resolved. In the Figures 13 and 14 both, the real and the imaginary part of the wavenumber are shown as functions of the frequency for an axisymmetric ($m = 0$) and an azimuthal mode ($m = 1$), respectively. These modes have some properties in common with hydrodynamic instability modes: they travel in the direction of flow (except for a certain frequency interval for $m = 1$) while their amplitudes grow, and they exist only with superimposed mean flow, i. e., with $\overline{U} \to 0 \;\Rightarrow\; \alpha \to \infty$ which however is found to be valid only at low frequencies. In the respective eigenvalue problem, uniform compressible mean flow in a circular pipe has been considered. The acoustic admittance of the wall can be reliably computed from the geometrical parameters of the resonator cavities [29], and the jump condition (11) has been used to fix the boundary condition.

Some data points obtained from the pressure distribution in the lined duct section (Sect. 2.1.2 and Ref. [29]) are displayed in Fig. 13, and the respective theoretical values are correspondingly marked. While, with $\overline{U}/c = \mathrm{const} = 0.23$, the measured phase constant $\Re\{\alpha\}$ strongly increases as a function of the frequency and the spatial growth rate $-\Im\{\alpha\}$ exhibits a clear maximum at 1077 Hz, the theoretical values only slightly depend on the frequency above the resonance frequency (840 Hz), and even worse, $\Re\{\alpha\}$ has the opposite slope. When in contrast the flow velocity is varied and the data points are taken at the frequencies of maximum spatial growth of the instability wave (Fig. 6), then the general trend is reproduced for $\Re\{\alpha\}$ but not for $\Im\{\alpha\}$. Also the dispersion relations shown in Fig. 14 for $m = 1$ are only marginally compatible with the observations. After all, the wavenumber depends nearly exclusively on $\omega/\overline{U}$ at low frequencies, and this kind of dependency has also been observed with the spectral peaks and the associated coherence of the pressure fluctuations in the backmost cavity (see the peak marked by C in Fig. 10). The theoretical wavenumbers at the respective peak frequencies are marked in Fig. 14

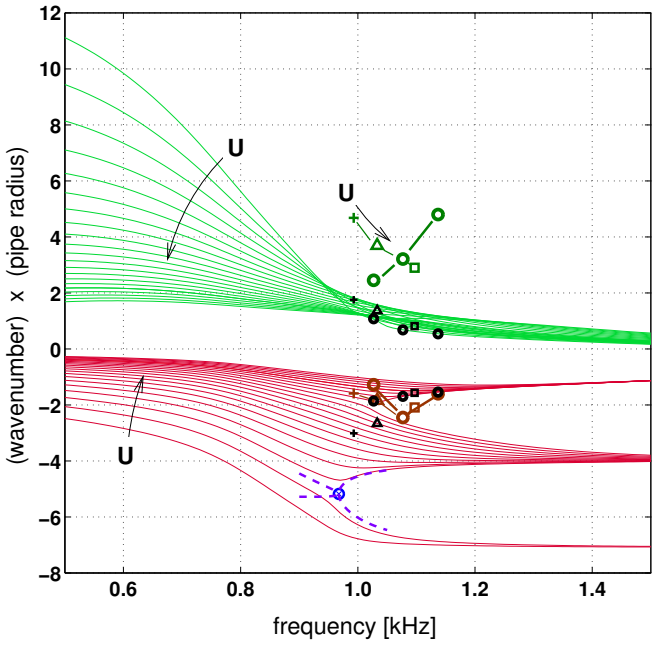


Figure 13. Dispersion relations of a $m = 0$ mode in a lined duct with dimensions according to Fig. 1. The flow velocity of the uniform compressible mean flow has been varied between $0.1 \leq \overline{U}/c \leq 0.3$ in steps of 0.01. The real (green) and the imaginary (red) parts of the wavenumber are shown in the upper and the lower part of the Figure, respectively. The experimental data (colored, from Ref. [29]) and the corresponding theoretical values (black) refer to $\overline{U}/c = 0.17(+)$, $0.20(\triangle)$, $0.23(\circ)$, $0.25(\square)$. The (blue) circled cross and the dashed lines describe a bifurcation (see text).

by circles which happen to appear at equivalent points of the dispersion relations, characterized e. g. by the maximum of $\Re\{\alpha\}$. Likewise it can only be speculated that the structure found in the theoretical dispersion relations at frequencies around 1000 Hz might be related to the increase of the pressure drop that is achieved by the excitation of the antisymmetric mode at similar frequencies (see Fig. 12).

The instability denoted by C in Fig. 10 and marked by the circles in Fig. 14 obviously does not rely on the compressibility of the air. The dynamics behind this instability is very similar to the Kelvin-Helmholtz instability which results from a balance of the accelerating forces that the fluid elements experience on both the sides of shear layers due to the unsteady deflection of the shear layer. In the present case the fluid flow in the interior of the duct behaves like an oscillating jet which forces the surrounding fluid in the cavities to oscillate in the azimuthal direction. So the accelerating forces are proportional to $\hat{\eta}_0 \omega^2$ in the cavities and proportional to $\hat{\eta}_U(\omega - \alpha U)^2$ in the flow. The sum of these forces has to cancel out which yields $\alpha = (\omega/U)(1 \mp i\kappa_\eta)$ wherein $\kappa_\eta = \sqrt{\hat{\eta}_0/\hat{\eta}_U}$ depends on the geometric details of the

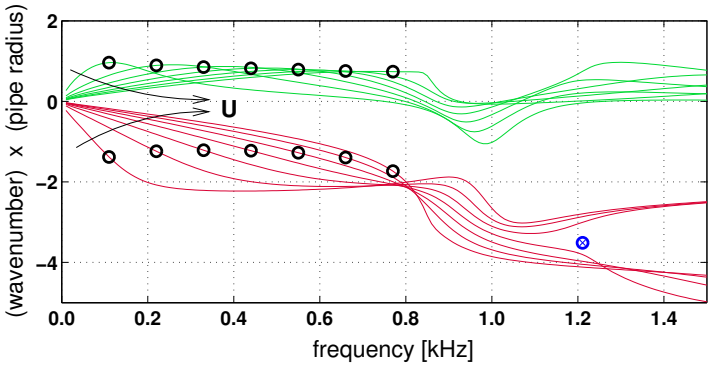


Figure 14. Like Fig. 13, but for the $m = 1$ mode, $0.05 \leq \overline{U}/c \leq 0.35$ in steps of 0.05. The circles mark the frequencies that correspond to the peak C in Fig. 10(a).

fluid oscillation. Comparing this result with the dispersion relations in Fig. 14 yields $\kappa_\eta = 1$ in the low frequency limit.

4.1.3 Bifurcations and absolute instabilities

Besides the discrepancies between the theoretical and experimental results the dispersion relations exhibit unexpected branch points which particularly show up with $\Im\{\alpha\}$. Two of these are marked by circled crosses in Figs. 13 and 14. The existence of such branch points has serious consequences for the stability of the flow because they may be associated with absolute instabilities. In homogeneous, absolutely unstable flow a locally excited perturbation temporally grows and finally spreads over the whole flow region. By contrast, a convective instability which is supposed to cause the spatial growth of the pressure fluctuations in the resonator section, results in amplifying perturbations that propagate away from the source; so after the source has been switched off the perturbation decays at fixed locations.

Whether an absolute instability exists is commonly investigated by means of the group velocity $\mathrm{d}\Omega_n/\mathrm{d}\alpha$. If for any mode n a complex wavenumber α_{vgr0} is found such that $\mathrm{d}\Omega_n/\mathrm{d}\alpha = 0$ for $\alpha = \alpha_{\mathrm{vgr0}}$, a wave packet with the complex mid-frequency $\omega_{\mathrm{vgr0}} = \Omega_n(\alpha_{\mathrm{vgr0}})$ will stay where it is. It then depends on the sign of $\Im\{\omega_{\mathrm{vgr0}}\}$ whether the amplitude of this in principle infinitely extended wave package grows ($\Im\{\omega_{\mathrm{vgr0}}\} > 0 \quad \Rightarrow$ absolute instability) or decays. It is to be annotated that the group velocity, particularly if it is not real, does not allow any conclusion about the direction of 'propagation' of the mode, i. e. about the question, whether the mode contributes to the wave field upstream or downstream of the source. This so called causality problem has to be resolved (Sect. 4.1.4) if no absolute instability is encountered and if a given mode shall be checked for convective instability. This is the case if the mode grows in the causality direction.

The relation between absolute instability and the existence of a branch point in the dispersion relation becomes obvious when the dispersion relation is expanded in

powers of $\delta\alpha = \alpha - \alpha_{\mathrm{vgr0}}$. With $\delta\omega = \omega - \omega_{\mathrm{vgr0}}$ one obtains

$$\begin{aligned} \Omega_n(\alpha_{\mathrm{vgr0}} + \delta\alpha) &= \omega_{\mathrm{vgr0}} + a_2\delta\alpha^2 + \mathcal{O}(\delta\alpha^3) \quad \text{with} \quad a_2 = \frac{1}{2}\frac{\mathrm{d}^2\Omega_n(\alpha_{\mathrm{vgr0}})}{\mathrm{d}\alpha^2} \\ \Rightarrow \quad \delta\omega &= a_2\delta\alpha^2 + \mathcal{O}(\delta\alpha^3)\,. \end{aligned} \tag{12}$$

Hence the inverse function $\alpha = \mathrm{inv}\{\Omega_n\}(\omega)$ has two branches in the neighborhood of ω_{vgr0}, namely $\delta\alpha(\omega) = \pm(\delta\omega/a_2)^{1/2}$ which have a common point if and only if ω_{vgr0} is exactly on the path through the complex ω-plane. Otherwise the two branches are supposed to be the dispersion relations of two different modes. These however are mixed when a parameter of the eigenvalue problem is continuously varied such that ω_{vgr0} crosses over the path through the ω-plane (also the path through the ω-plane may be subject to deformation); then the four parts of branches that emanate from the branch point are differently combined before and after the crossing. In the case considered in Fig. 13 the path through the ω-plane is the real axis and the varied parameter is the flow velocity. So, with $\Im\{\omega_{\mathrm{vgr0}}\} = 0$ the flow is marginally stable for $\overline{U}/c = 0.1155$ but theoretically becomes absolutely unstable ($\Im\{\omega_{\mathrm{vgr0}}\} > 0$) with $\overline{U}/c > 0.1155$. The dashed curves in Fig. 13 illustrate the four possible dependencies of $\Im\{\alpha\}$ on the frequency: each of the two low-frequency branches can combine with each of the two high-frequency branches. Only two of these possibilities are shown with the dispersion relations plotted in the Figure. The absolute instability that is expected above a critical flow velocity has not been observed in the experiments, and before we further search for indications of an absolute instability we consider the convective instabilities and the causality problem.

4.1.4 Convective instability and causality problem

As mentioned earlier the direction of spatial growth $-\Im\{\alpha\}$ of a mode has to agree with its causality direction $\nu_{\mathrm{caus}}(= \pm 1)$, i. e. $\nu_{\mathrm{caus}} \cdot \Im\{\alpha\} < 0$, if the mode is to be classified as convective unstable. A trivial but laborious way to determine ν_{caus} is the computation of the total field which is caused by a local source switched on at $t = 0$. To avoid this effort Briggs [37] has developed a criterion which allows to determine the causality direction on the basis of the dispersion relation. He replaces the switching-on by exponential growth of the source amplitude, and the imitation of the discontinuous switching-on is the better the faster the amplitude grows, i. e. if $\Im\{\omega\} \to \infty$. It is to be expected that the excited field then concentrates at the location of the source because no time for the spreading of the field has been left. So each mode decays in its direction of causality: $\nu_{\mathrm{caus}} \cdot \Im\{\alpha\} > 0$. Now, according to the Briggs criterion, $\Im\{\omega\}$ is gradually reduced to zero and $\nu_{\mathrm{caus}} \cdot \Im\{\alpha\}$ is checked for changing the sign which indicates that the respective mode finally (at real frequencies) grows in the causality direction and thus is subject to convective instability.
However two conditions have to be fulfilled before the Briggs criterion can be applied:
(i) The result must be independent of the path through the ω-plane on which $\Im\{\omega\}$ is reduced to zero.
(ii) An upper bound ω''_{max} must exist such that $\nu_{\mathrm{caus}} \cdot \Im\{\alpha\} > 0$ for all ω with $\Im\{\omega\} > \omega''_{\mathrm{max}}$; an equivalent and actually verifiable condition is the existence of an upper bound ω''_{max} such that $\Im\{\Omega_n(\alpha)\} \le \omega''_{\mathrm{max}}$ for all real α.

The condition (ii) is frequently violated, in particular, if free shear layers with finite thickness are approximated by infinitesimally thin vortex layers (then in many cases a hydrodynamic mode is encountered with $\Im\{\Omega(\alpha)\} \to \infty$ for $\alpha \to \infty + 0i$. This problem has been investigated by Jones and Morgan [38] and later by Crighton and Leppington [39], and the Briggs criterion has been modified by the latter authors. They propose to trace $\alpha = \mathrm{inv}\{\Omega_n\}(\omega)$ while running on a quarter arc of a circle in the first quadrant of the ω-plane, i. e. from $i|\omega|$ to $|\omega|$ with $|\omega|$ kept constant. Rienstra and Peake [16] have compared these two criteria for the wave propagation in resonatingly lined circular ducts which are comparable to our case, except for the much higher quality factor of our resonators. The authors find that the causality direction indeed differs between the two criteria in several cases. However the authors do not discuss the existence of branch points of the dispersion relations which cause a violation of the condition (i). We find a great number of branch points with $\Im\{\omega_{\mathrm{vgr}0}\} > 0$ not depending on whether a flat velocity profile or a more realistic representation of the turbulent flow in the lined duct is assumed [35]. We even suspect that all the modes in the forth quadrant of the α-plane are connected via branch points.

4.1.5 Search for experimental indications of absolute instability

The existence of branch points with $\Im\{\omega_{\mathrm{vgr}0}\} > 0$ in the theoretical dispersion relations raises the question whether there are experimental observations that hint at absolute instabilities of the flow and have been overlooked up to now. In most cases a steady, however absolutely unstable flow assumes a spatially and temporally periodic state the frequency and wavenumber of which is not far from $\Re\{\omega_{\mathrm{vgr}0}\}$ and $\Re\{\alpha_{\mathrm{vgr}0}\}$ – a salient example is the vortex street in the wakes of bluff bodies [40]. An axisymmetric oscillatory state in the resonator section should radiate sound into the rigid pipes connected to the resonator section, and the distribution of the radiated pressure amplitudes should be similar to the distribution of a sine wave; by contrast turbulent fluctuations that have passed a narrow-band filter or a narrow-band amplifier are normally distributed.

Brandes [29,30] indeed finds narrow peaks in the pressure spectra which are radiated into the pipe upstream of the lined duct section. While the peak denoted by D in Fig. 2(c) reflects only a weak example of this kind of instability, much stronger peaks have been found with higher resonance frequencies (2934 Hz in most of Brandes' experiments). The signals that result from this instability have all the mentioned properties of self-excited oscillations. However the frequency of the oscillation is inversely proportional to the the spacing of the cavities. So it is very likely that the oscillations are generated by a feedback loop rather than by an absolute instability. It is remarkable that also narrow-band sound amplification is observed at the frequencies of these oscillations; the amplified sound is radiated in the direction against the flow [25], in contrast to the sound amplification shown in Fig. 4. Brandes explains this kind of sound amplification by partial synchronization of the self-exited oscillation with the incident sound. Thereby the amplitude of the synchronized sound radiation is proportional to the much smaller amplitude of the incident sound wave. However with high incident sound amplitudes, the self-excited oscillation is completely synchronized, and it is even possible to considerably change its frequency. Yet, neither

an increase of the total radiated sound power nor an increase of the static pressure drop is effected by the synchronization.

The sound pressure that originates from the mainly studied resonator section and that causes the spectral peak A^1 in Fig. 2 exhibits some properties which might be due to an absolute instability and which have already been mentioned in connection with Fig. 3. With all the studied configurations consisting of at least six of the cavities shown in Fig. 1 a particular flow velocity $\overline{U}_{\mathrm{jump}}$ is found at which the peak frequency jumps from ca. 1200 Hz to ca. 1260 Hz. At somewhat lower flow velocities the peak width decreases with increasing flow velocity while the increase of the peak amplitude is steeper than at other flow velocities. In this range of flow velocities also a strong change of the distribution of the pressure amplitudes is found which finally is very similar to the distribution of a sine wave. In addition the pressure oscillation can be synchronized with incident sound [35]. So the acoustic transmission coefficient becomes particularly large at these flow velocities (see Fig. 4 for 16 cavities with $\overline{U}_{\mathrm{jump}}/c \approx 0.35$). However while $\overline{U}_{\mathrm{jump}}$ strongly depends on the number of cavities and on the geometry of the duct cross-section the 'jump'-frequency does not. Furthermore the strong oscillation of the pressure becomes weaker or even breaks off when the flow velocity exceeds $\overline{U}_{\mathrm{jump}}$. These latter facts do not belong to the obvious concomitants of an absolute instability.

4.2 Effect of shear stress

4.2.1 Turbulent Stokes layer and its effect on the boundary condition at the wall

The essential simplifying assumptions which have been made in the previously discussed calculations are, except for the axial homogeneity of the mean flow, (i) uniform mean flow over the cross-section, (ii) homogeneous locally reacting compliance of the wall, and (iii) disregarding of shear stress. (i) The introduction of an adequate approximation of the real velocity profile has indeed an appreciable effect on the dispersion relations (not shown here), however the general dependency on the frequency and on the flow velocity remains unchanged. (ii) The wavelength of the observed instability wave is much larger than the spacing of the cavities which makes it unlikely that the inhomogeneity of the wall plays an essential role in the phenomena to be explained here. In fact the sound amplification has been observed to increase when the ratio between the width of the cavities and the wavelength of the instability wave is decreased [26]. So the inhomogeneity is obviously not a precondition of the considered instability. (iii) Finally the possible effect of shear stress remains to be studied more extensively.

In the flow to be considered the turbulent mixing and the generation of Reynolds stress are still effective very close to the wall here because the permeability of the wall allows the turbulent eddies to pass through the wall. So the axial momentum of the eddies is very efficiently transported to the wall which means that the Reynolds shear stress $\overline{\tau}'' = -\rho\overline{u'v'}$ including the wall shear stress $\overline{\tau}_{\mathrm{w}}$ are much higher (by nearly one order of magnitude) than with an impervious smooth wall. Nevertheless the actual transfer to the wall is effected by viscosity within a very thin layer. Within this layer the mean shear rate $\mathrm{d}U/\mathrm{d}y = \overline{\tau}_{\mathrm{w}}/\mu$ must be extremely high so that the flow

velocity increases from zero to a value U_ϵ, nearly discontinuously; $U_\epsilon \approx \overline{U}/4$ has been estimated on the basis of the sound reflection from and the transmission through the resonator section measured at frequencies far below the resonance frequency [19].

Thus the wall-normal flow velocity $\widetilde{v}_\mathrm{w}$ of the instability wave imposes an oscillating component $\widetilde{\tau}_\mathrm{w} = -\rho U_\epsilon \widetilde{v}_\mathrm{w}$ on the wall-normal transport of axial momentum; $\widetilde{v}_\mathrm{w} = \langle v_\mathrm{w} \cdot \exp(i\omega t)\rangle \cdot \exp(-i\omega t)$ is the phase average by which the coherent part at the frequency ω is filtered out from the fluctuating quantity $v_\mathrm{w}(t)$. The considered oscillation $\widetilde{\tau}_\mathrm{w}$ of the wall shear stress is by orders of magnitude greater than with an impervious smooth wall even if the considerable increase of $\widetilde{\tau}_\mathrm{w}$ by the turbulent mixing [41] is taken into account. The transfer of axial momentum to the wall by the wall-normal particle velocity, and the associated high acoustic shear stress have scarcely been regarded in aero-acoustics so far. This mechanism is observed also with openings subject to grazing flow [42], and it can be used with deflected mean flow to arbitrarily increase the quality factor of resonators [43–45]. Pöthke and Ronneberger [42] determine the amplitude of the shear force at transverse slits in the wall of the flow duct; these had the same dimensions as openings of the cavities in the present experiments. At low frequencies (compared to the ratio between flow velocity and slit width) the same value of the shear force is obtained as in the aforementioned investigation by Rebel and Ronneberger [19], but the shear force amplitude increases considerably as a function of the frequency according to Ref. [42]. Yet, a frequency-independent value $U_\epsilon = \overline{U}/4$ is assumed in the following estimates.

The oscillation of the wall shear stress in the lined duct section excites a shear wave which is propagated by the wall-normal gradient of the shear stress $\widetilde{\tau} = \widetilde{\tau}'' + \widetilde{\tau}_\mu$ into the flow region. The viscous part $\widetilde{\tau}_\mu = \mu \cdot \partial\widetilde{u}/\partial y$ of the shear stress is small compared to the turbulent part and therefore will be disregarded. The layer within which the shear wave decays is called 'turbulent Stokes layer' following the denomination in the laminar case. Though the thickness δ_S of the Stokes layer is much greater in the turbulent case than in the laminar case we assume that δ_S is still small compared to the lateral extent of the field of the instability wave which decays approximately like $\exp(-|\Re\{\alpha\}|y)$. So we attempt to subsume the influence of the shear stress under the boundary condition at the wall.

In order to study the influence of the shear stress on the deflection of the streamlines, the relation (10) is introduced into the equations of conservation of mass and momentum. For the sake of clearness a 2D flow is assumed composed of the mean flow and the field of a coherent oscillation while the effect of the turbulent fluctuations are already taken into account by the Reynolds shear stress. Then one obtains for the deflection of the fictitious streamlines

$$\overline{\rho}\frac{\mathrm{D}^2}{\mathrm{D}t^2}\frac{\partial\widetilde{\eta}}{\partial y} + \frac{1}{c^2}\frac{\mathrm{D}^2\widetilde{p}}{\mathrm{D}t^2} - \frac{\partial}{\partial x}\left(\frac{\partial\widetilde{p}}{\partial x} - \frac{\partial\widetilde{\tau}_{xx}}{\partial x} - \frac{\partial\widetilde{\tau}_{xy}}{\partial y}\right) = 0\,. \tag{13}$$

The terms $\widetilde{\tau}_{xx}$, $\widetilde{\tau}_{xy} = \widetilde{\tau}'' + \widetilde{\tau}_\mu$ (and $\widetilde{\tau}_{yy}$) are the components of the tensor which comprises the oscillatory parts of the viscous and the Reynolds stresses. A clearer version of Eq. (13) is obtained when a single mode is considered such that $\partial/\partial x$ and $\mathrm{D}/\mathrm{D}t$ can be replaced by $i\alpha$ and $-i\breve{\omega}$, respectively:

$$\frac{\mathrm{d}\hat{\eta}}{\mathrm{d}y} = -\frac{\hat{p}}{\overline{\rho}c^2} - i\alpha\frac{i\alpha\hat{p} - i\alpha\hat{\tau}_{xx} - \mathrm{d}\hat{\tau}_{xy}/\mathrm{d}y}{\overline{\rho}\breve{\omega}^2}\,. \tag{14}$$

The term $\mathrm{d}\hat{\eta}/\mathrm{d}y$ is the oscillation of the wall-normal spacing of the streamlines or in other words the oscillation of the thickness of the streamtubes. $\mathrm{d}\hat{\eta}/\mathrm{d}y$ is effected by the compressibility $(\overline{\rho}c^2)^{-1}$ of the fluid and by the oscillation of the length of the elements of the streamtubes (second term on the right-hand side of Eq. (14)). The latter is described by the responsible accelerating forces. The terms which depend on the pressure will be omitted from the further consideration since these terms are already included in the differential equations which describe the dynamics of the fluid due to the action of the pressure (e. g., the Rayleigh equation (4) with the inclusion of the compressibility effects). This includes that the velocity field is assumed to be composed of two non-interacting parts which are exclusively governed by the pressure and by the shear stress, respectively.

The oscillation of the Stokes layer thickness is computed by the integration of $\mathrm{d}\hat{\eta}/\mathrm{d}y$ through the Stokes layer. Complying with the assumption that the Stokes thickness δ_S is much smaller than the lateral extent $|\Re\{\alpha\}|^{-1}$ of the wave field only the limit $|\Re\{\alpha\}|\delta_\mathrm{S} \to 0$ will be considered. The shear stress is assumed to decay from its value $\hat{\tau}_\mathrm{w}$ at the wall to zero at the outer edge of the Stokes layer according to $\hat{\tau}_{xy}(y) = \hat{\tau}_\mathrm{w}\cdot\hat{\theta}(\check{y})$ with $\check{y} = y/\delta_\mathrm{S}$. Then only the term with $\mathrm{d}\hat{\tau}_{xy}/\mathrm{d}y = (\hat{\tau}_\mathrm{w}/\delta_\mathrm{S})(\mathrm{d}\hat{\theta}/\mathrm{d}\check{y})$ contributes to the integral and since this term is zero outside the Stokes layer the integration can be extended to infinity without changing the result:

$$\hat{\delta}_\mathrm{S} := \hat{\eta}_\tau(\delta_\mathrm{S}) - \hat{\eta}_\tau(0) = \int_0^{\delta_\mathrm{S}} \frac{\mathrm{d}\hat{\eta}_\tau}{\mathrm{d}y}\mathrm{d}y = \frac{i\alpha\hat{\tau}_\mathrm{w}}{\omega^2\,\overline{\rho}} \int_0^{\infty} \frac{\mathrm{d}\hat{\theta}/\mathrm{d}\check{y}}{[1-\alpha U(\check{y})/\omega]^2}\mathrm{d}\check{y}\,. \tag{15}$$

Thus, the amplitude $\hat{\delta}_\mathrm{S}$ of the oscillation of the Stokes layer thickness is proportional to the amplitude of the wall shear stress $\hat{\tau}_\mathrm{w} = -\rho U_\epsilon \hat{v}_\mathrm{w} = i\,\omega\rho U_\epsilon \hat{\delta}_\mathrm{w}$, and consequently is proportional to the amplitude $\hat{\delta}_\mathrm{w} := \hat{v}/i\omega|_{y=-0}$ of the fictitious deflection of the wall. The deflection of the streamlines at the border of the Stokes layer is then given by $\hat{\delta}_\mathrm{w} + \hat{\delta}_\mathrm{S}$. So the ratio $|\hat{\delta}_\mathrm{S}|/|\hat{\delta}_\mathrm{w}|$ can be used to depict the influence of the shear stress on the boundary condition at the wall.

In order to estimate $\hat{\theta}(\check{y})$, and with it the integral on the right-hand side of Eq. (15), we follow the common assumption that the relation between the Reynolds stress tensor and the respective flow field is local and can be described by the so called eddy viscosity μ_t. For the mean flow through a circular pipe the wall of which is rough and is assumed to be a fair equivalent to the pervious wall [19], one finds $\overline{\tau}''/(\mathrm{d}\overline{u}/\mathrm{d}y) =: \overline{\mu}_\mathrm{t} = \rho U_\mu \cdot (y + y_0)$, wherein $U_\mu = a_u\overline{U}$ and $y_0 = a_y R$ (here $a_u = 0.044$, $a_y = 0.017$) are fitted to the mean wall shear stress $\overline{\tau}_\mathrm{w}$ and the average mean flow velocity $\overline{U}$. Unfortunately much less is known about the relation between $\widetilde{\tau}''$ and $\widetilde{\boldsymbol{u}}$. Hartmann [46] has investigated the propagation of shear waves in a turbulent channel flow at low Reynolds numbers by means of a direct numerical simulation which he validated on the basis of comprehensive experimental data [47–51]. In accordance with theoretical considerations he finds that the turbulence is equivalent to a viscoelastic fluid with regard to the response of the shear stress to a shearing deformation. However, the extrapolation of his results to the present Reynolds numbers and frequencies yields that the time constant that characterizes the memory of the turbulence is much shorter than the periods of the considered oscillations. So the ratio

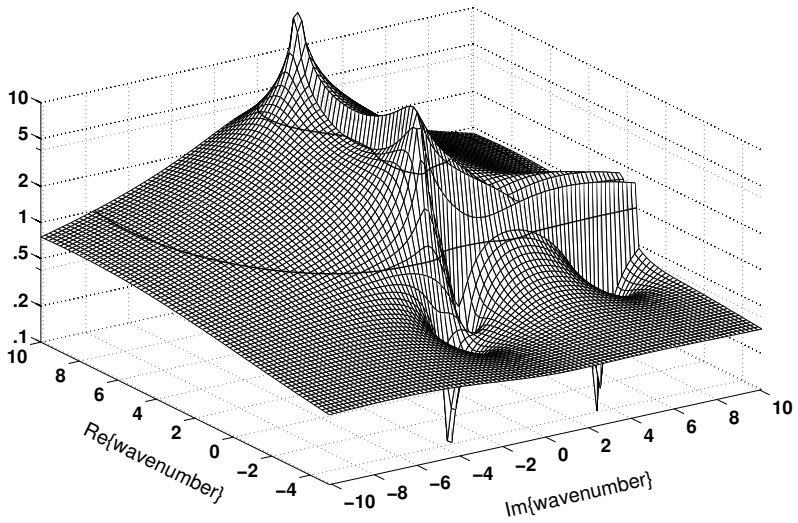


Figure 15. The amplitude ratio $|\hat{\delta}_\mathrm{S}|/|\hat{\delta}_\mathrm{w}|$ of the thickness of the Stokes layer and the deflection of the wall with contour lines at $|\hat{\delta}_\mathrm{S}|/|\hat{\delta}_\mathrm{w}| = 1$ and $|\hat{\delta}_\mathrm{S}|/|\hat{\delta}_\mathrm{w}| = 5$ is plotted as a function of the complex wavenumber of the modes that propagate in the lined duct. The wavenumber is normalized to the duct radius. Parameters: $f = 1100\,\mathrm{Hz}$ and $\overline{U}/c = 0.25$.

$\tilde{\mu}_\mathrm{t} = \hat{\tau}''/(\mathrm{d}\hat{u}/\mathrm{d}y)$ should be real and should assume the value $\tilde{\mu}_\mathrm{t} \approx 2\overline{\mu}_\mathrm{t}$ according to the somewhat uncertain extrapolation of Hartmann's results. In a WKBJ approximation the wavenumber $\gamma = (-i\omega\rho/\tilde{\mu}_\mathrm{t})^{1/2}$ of shear waves in homogeneous fluids is transferred to the actual inhomogeneous case, yielding $\hat{u}_\tau(y) \approx \hat{u}_\tau(0)\exp(\int_0^y \gamma(y')\mathrm{d}y')$. So Eq. (15) can be evaluated with $\hat{\theta}(y) \approx (\tilde{\mu}_\mathrm{t}\gamma\hat{u}_\tau)(y)/(\tilde{\mu}_\mathrm{t}\gamma\hat{u}_\tau)_\mathrm{w}$.

The ratio $|\hat{\delta}_\mathrm{S}|/|\hat{\delta}_\mathrm{w}|$ for a typical combination of parameters ($1100\,\mathrm{Hz}$, $\overline{U}/c = 0.25$) is depicted in Fig. 15 as a function of the wavenumber α. For wavenumbers that have been observed in the experiments (see Fig. 13), $|\hat{\delta}_\mathrm{S}|/|\hat{\delta}_\mathrm{w}|$ assumes values which are typically much greater than unity. So obviously the oscillation of the Stokes layer thickness turns out to be the governing factor in the boundary condition at the wall.

4.2.2 Influence of the oscillation of the Stokes layer thickness on the dispersion relations and open questions

First computational results reveal that the dispersion relations are indeed strongly affected by the oscillation of the Stokes layer thickness. Similar to the experimental observations one of these dispersion relations (not shown here) exhibits maxima of the spatial growth rate ($-\Im\{\alpha\}$), and at the same frequencies, a positive slope of the phase constant. However, this occurs at frequencies below the resonance frequency, and the phase constant is much too small and mostly even negative. Furthermore the dependency on the flow velocity is still contrary to the experiment, and the bifurcations have not disappeared from the dispersion relations.

Nevertheless, with regard to the approximations that have been made in the previous section, it is still open to question whether the gap between experiment and theory can be closed by taking the shear stress into account. In fact more than one of these assumptions turn out to be problematic. Even the very first assumption that the velocity can be decomposed into two non-interacting parts which are determined by the pressure and by the shear stress, respectively, cannot be maintained with low phase velocities and small ratios $\Im\{\alpha\}/\Re\{\alpha\}$. In such cases a so called critical layer y_c exists with $\breve{\omega}(y_c) \approx 0$ so that the right-hand side of the Rayleigh equation (4) as well as the integrand in Eq. (15) are nearly singular there. This unphysical singularity causes, among others, the discontinuity of $|\hat{\delta}_S|/|\hat{\delta}_w|$ when the real α-axis is passed (see Fig. 15). In reality the shear stress adjusts itself such that the numerator in Eq. (14) becomes small together with $\breve{\omega}$. Also is the Stokes layer thickness not small, i. e. $|\Re\{\alpha\}|\delta_S = \mathcal{O}(1)$ with the wavenumbers observed in the experiment nor is the WKBJ approximation unquestionable, close to the wall.

So it appears to be necessary to solve the full problem before a final answer to the question can be expected whether or not the observed instability waves and their dependency on the essential parameters can be described by the wave propagation in a homogeneous environment (i. e. that all parameters are assumed to be independent of x). This leads to a fourth order differential equation if eddy-viscosity is assumed, and the algorithm for the solution of the respective eigenvalue problem has to be enhanced for this purpose. Yet it is questionable whether the response of the turbulence can indeed be described by an eddy-viscosity in view of the fact that the turbulence is by itself subject to flow instability so that its state can be anticipated to be far from equilibrium. Beyond that, regarding the bifurcations of the dispersion relations and the associated absolute instabilities, it is suspected that it is indispensable to take the spatial development of the flow into account. This might lead to the conclusion that the distance within which the flow is subject to a certain absolute instability is too short for the development of a global instability [52], however that the wave propagation is nevertheless governed by the absolute instability.

5 Summary and concluding remarks

A strong convective instability has been found with a short section of a flow duct which was provided with a resonating lining. Sound waves that propagate in the direction of the mean flow as well as turbulent pressure fluctuations are amplified by this instability, and the pressure drop along the lined duct section is drastically increased due to the power consumption by the excited amplifying instability wave. While the sound radiation from the rear end of the lined duct section into the trailing rigid pipe is possible only with the fundamental axially symmetric mode ($m = 0$), the increase of the pressure drop is achieved also with an antisymmetric mode ($m = 1$) nearly as efficiently as with the ($m = 0$) mode. This is in favour of technical applications of the phenomenon because the mostly undesired radiation of the controlling sound is avoided in this way. One such application, namely the active suppression of low-frequency sound transmission through the lined duct section, has already been demonstrated. The sound-induced pressure gradient can be increased by reducing

the free cross-section of the flow duct by a cylindrical coaxial body such that only an annular gap remains open to the flow. A sound-induced increase of the pressure gradient of up to 5% of the dynamic pressure per gap width has been observed, however the relative increase of the pressure drop remains rather independent of the gap width.

Many of the observed phenomena can be explained, at least on a qualitative level, if the convective instability of the flow is taken for granted. The enhancement of the Reynolds shear stress caused by the spatial growth of the instability wave plays an essential role in the interaction between the mean and the oscillating part of the flow. Thus both, the mean velocity profile as well as the intensity and the structure of the turbulence experience a strong alteration along the way through the lined duct section, and the spatial development of the flow on its part strongly affects the growth of the instability wave. The exact nature of the instability however remains unclear. So far, the search for an adequate description of the dynamics of the flow has been restricted to approaches that are based on homogeneous mean flow conditions or at the most on a slowly developing mean flow, and on a mode decomposition of the oscillating field. The interaction between the instability wave and the turbulence has been introduced by means of an eddy-viscosity and has been found to have a dominant influence on the dynamics of the flow. All these attempts have ended up in unrealistic dispersion relations and particularly in the detection of absolute instability which however has not been observed in the experiments. Therefore, though not all the implications of the eddy-viscosity have been studied up to now, we conjecture that the spatial development of both, the mean flow and the instability wave have to be considered in a joint analysis.

A few of the observations have not yet been explained even on the qualitative level mentioned above. Among these are the reduction (in contrast to an increase) of the pressure drop by superimposed sound as well as the frequency gap that exists between 1200 Hz and 1260 Hz for the $m = 0$ mode but is particularly favourable for the $m = 1$ mode; by the way a similar gap has been observed for the C-mode ($m = 1$) between 300 Hz and 380 Hz (see Fig. 10). Presumably these observations are caused by some interaction between different instability waves, whether turbulent or coherent, and can be explained only by deeper insight into the dynamics of the considered flow.

Acknowledgements. The authors are indebted to Sabine Förster, Jörg Rebel, Michael Krause, Lars Enghardt, Andreas Pöthke, Michael Brandes, Björn Lange, and Jakob Großer who, by their diploma- and doctoral theses have essentially contributed to the material which has been presented here. Part of this research has been funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) whose support is gratefully acknowledged.

References

[1] M. Lighthill, 'On Sound Generated Aerodynamically. I. General Theory', Proc. Roy. Soc. Lond. Ser. A **211**, 564 (1952).

[2] E. Meyer, F. Mechel, and G. Kurtze, 'Experiments on the Influence of Flow on Sound Attenuation in Absorbing Ducts', J. Acoust. Soc. Am. **30**, 165 (1958).

[3] F. Mechel, 'Schalldämpfung und Schallverstärkung in Luftströmungen durch absorbierend ausgekleidete Kanäle', Acustica **10**, 133 (1960).

[4] J. Bruggeman, A. Hirschberg, M. van Dongen, A. Wijnands, and J. Gorter, 'Flow induced pulsations in gas transport systems: analysis of the influence of closed side branches', J. Fluids Eng. **111**, 484 (1991).
[5] J. Gray, 'Studies in Animal Locomotion : VI. The Propulsive Powers of the Dolphin', J. Exp. Bio. **13**, 192 (1936).
[6] M. O. Kramer, 'Boundary layer stabilization by distributed damping', J. Am. Soc. Nav. Eng., **74**, 341 (1960).
[7] T. Benjamin, 'Effects of a Flexible Boundary on Hydrodynamic Stability', J. Fluid Mech. **9**, 513 (1960).
[8] M. Gad-el Hak, 'Compliant coatings: a decade of progress', Appl. Mech. Rev. **49**, 147 (1996).
[9] P. Carpenter, C. Davies, and A. Lucey, 'Hydrodynamics and compliant walls: Does the dolphin have a secret?', Current Science **79**, 758 (2000).
[10] B. Tester, 'The Propagation and Attenuation of Sound in Ducts Containing Uniform 'Plug' Flow', J. Sound Vib. **28**, 151 (1973).
[11] B. Nilsson and O. Brander, 'The Propagation of Sound in Cylindrical Ducts with Mean Flow and Bulk-reacting Lining I. Modes in an Infinite Duct', IMA J. Appl. Math. **26**, 269 (1980).
[12] W. Möhring and W. Eversman, 'Conversion of acoustic energy by lossless liners', J. Sound Vib. **82**, 371 (1982).
[13] W. Koch and W. Möhring, 'Eigensolutions for liners in uniform mean flow ducts', AIAA Journal **21**, 200 (1983).
[14] M. Quinn and M. Howe, 'On the production and absorption of sound by lossless liners in the presence of mean flow', J. Sound Vib. **97**, 1 (1984).
[15] S. Rienstra, 'Hydrodynamic Instabilities and Surface Waves in a Flow over an Impedance Wall', in *Proc. IUTAM Symp. 'Aero- and Hydroacoustics' 1985 Lyon* (Springer Verlag, Heidelberg, 1986), p. 483.
[16] S. Rienstra and N. Peake, 'Modal Scattering at the Impedance transition in a Lined Flow Duct', in *11th AIAA/CEAS Aeroacoustics Conf., 2005* (Amer. Inst. Aeronaut. Astronaut., 2005), paper 2005-2805.
[17] D. Crighton, 'The Kutta condition in unsteady flow', Annu. Rev. Fluid Mech. **17**, 411 (1985).
[18] D. Ronneberger and S. Förster, 'Einfluß der Schubspannung auf die Schallausbreitung in Strömungskanälen mit porösen Wänden', in *Fortschritte der Akustik, DAGA '87* (Deutsche Gesellschaft für Akustik, 1987), p. 361.
[19] J. Rebel and D. Ronneberger, 'The effect of shear stress on the propagation and scattering of sound in flow ducts', J. Sound Vib. **158**, 469 (1992).
[20] D. Ronneberger, 'Zur Kutta-Bedingung bei der Schallausbreitung durch durchströmte Querschschnittssprünge und Lochplatten', in *Fortschritte der Akustik, DAGA '89* (Deutsche Gesellschaft für Akustik, 1989), p. 499.
[21] R. Graf, *Instationäre Strömungsablösung und Wirbeltransport an der Hinterkante einer Splitterplatte*, Diploma thesis, University of Göttingen, Göttingen (1998).
[22] R. Graf and D. Ronneberger, 'Die instationäre Kutta-Bedingung an der Hinterkante einer Splitterplatte', in *Fortschritte der Akustik, DAGA '98* (Deutsche Gesellschaft für Akustik, 1998), p. 600.
[23] H. Tokuno, *Anregbarkeit der Kelvin-Helmholtz-Instabilität in der freien Scherströmung hinter einer scharfen Kante*, Dissertation, University of Göttingen, Göttingen (2004).
[24] S. Schmitz, *Instationäre Strömungsablösung an der Hinterkante einer Splitterplatte*, Diploma thesis, University of Göttingen, Göttingen (2004).
[25] M. Krause, *Schallausbreitung in Strömungskanälen mit kassettierter Wandauskleidung*,

Diploma thesis, University of Göttingen, Göttingen (1990).
[26] M. Brandes and D. Ronneberger, 'Sound amplification in flow ducts with a periodic sequence of resonators', in *First Joint CEAS/AIAA Joint Aeroacoustics Conference, Munich, Germany* (Deutsche Gesellschaft für Luft- und Raumfahrt, 1995), p. 893.
[27] B. Lange and D. Ronneberger, 'Active Noise Control by Use of an Aeroacoustic Instability', Acta Acoustica/Acustica **89**, 658 (2003).
[28] M. Jüschke and D. Ronneberger, 'Control of Pipe Flow Resistance by Sound', in *Proc. Joint Congr. CFA/DAGA '04* (Deutsche Gesellschaft für Akustik, 2004), p. 131.
[29] M. Brandes, *Schallverstärkung in Strömungsknälen mit resonanzartiger Wandauskleidung*, Dissertation, University of Göttingen, Göttingen (1997).
[30] M. Brandes and D. Ronneberger, 'Schallverstärkung durch Phasensynchronisation einer strömungsakustischen Instabilität', in *Fortschritte der Akustik, DAGA '97* (Deutsche Gesellschaft für Akustik, 1997), p. 212.
[31] M. Brandes, L. Enghardt, and D. Ronneberger, 'Optimierung der Meßorte und der Beschallung bei akustischen Messungen im Strömungskanal', in *Fortschritte der Akustik, DAGA '92* (Deutsche Gesellschaft für Akustik, 1992), p. 945.
[32] Y. Aurégan, M. Leroux, and V. Pagneux, 'Abnormal behavior of an acoustical liner with flow', in *Proceedings Forum Acusticum 2005 Budapest* (European Acoustics Associations, 2005), p. 793.
[33] M. Brandes and D. Ronneberger, 'Wechselwirkung zwischen Schall und Strömung an Resonanz-Absorbern', in *Fortschritte der Akustik, DAGA '96* (Deutsche Gesellschaft für Akustik, 1996), p. 110.
[34] J. Großer, *Modellbildung für die Schallverstärkung in nachgiebig ausgekleideten Strömungskanälen*, Dissertation, University of Göttingen, Göttingen (2003), note.
[35] M. Jüschke, *Akustische Beeinflussung einer Instabilität in Kanälen mit überströmten Resonatoren*, Dissertation, University of Göttingen, Göttingen (2007).
[36] J. Großer and D. Ronneberger, 'Auf dem Weg zu einem akustisch regelbaren Strömungswiderstand', in *Fortschritte der Akustik, DAGA '03* (Deutsche Gesellschaft für Akustik, 2003), p. 502.
[37] R. Briggs, *Electron-Stream Interaction with Plasmas* (MIT Press, Cambridge Massachusetts, 1964).
[38] D. Jones and J. Morgan, 'The instability of a vortex sheet on a subsonic stream under acoustic radiation', Proc. Camb. Phil. Soc **72**, 465 (1972).
[39] D. Crighton and F. Leppington, 'Radiation properties of the semi-infinite vortex sheet: the initial-value problem', J. Fluid Mech. **64**, 393 (1974).
[40] W. Koch, 'Local instability characteristics and frequency determination of self-excited wake flows', J. Sound Vib. **99**, 53 (1985).
[41] D. Ronneberger and C. Ahrens, 'Wall shear stress caused by small amplitude perturbations of turbulent boundary-layer flow: an experimental investigation', J. Fluid Mech. **83**, 433 (1977).
[42] A. Pöthke and D. Ronneberger, 'Akustische Eigenschaften von turbulent überströmten Querschlitzen', in *Fortschritte der Akustik, DAGA '95* (Deutsche Gesellschaft für Akustik, 1995), p. 535.
[43] H. Preckel and D. Ronneberger, 'Ausnutzung strömungsbedingter negativer Impdanzen bei der Konstruktion eines akustischen Filters hoher Güte', in *Fortschritte der Akustik, DAGA '94* (Deutsche Gesellschaft für Akustik, 1994), p. 665.
[44] U. Tapken and D. Ronneberger, 'Strömungsakustische Mechanismen an frontal angeströmten Resonatormündungen', in *Fortschritte der Akustik, DAGA '98* (Deutsche Gesellschaft für Akustik, 1998), p. 578.
[45] M. Jüschke and D. Ronneberger, 'Ein Resonanzschalldämpfer mit umgelenkter Gleich-

strömung', in *Fortschritte der Akustik, DAGA '02* (Deutsche Gesellschaft für Akustik, 2002), p. 304.
[46] M. Hartmann, *Zeitlich modulierte Statistik der periodisch gestörten turbulenten Kanalströmung*, Dissertation, University of Göttingen, Göttingen (2001).
[47] M. Beykirch, *Untersuchung räumlich-zeitlicher Geschwindigkeitskorrelationen in einer periodisch gestörten turbulenten Kanalströmung*, Dissertation, University of Göttingen, Göttingen (1996).
[48] H. Flötke, *Modulation von Häufigkeitsverteilungen in der turbulenten Grenzschicht an einer schwingenen Wand*, Dissertation, University of Göttingen, Göttingen (1987).
[49] K. Breuer, *Messung der Schubspannung in einer zeitlich modulierten turbulenten Kanalströmung*, Diploma thesis, University of Göttingen, Göttingen (1995).
[50] K. Hesse, *Zeitabhängigkeit der spektralen Eigenschaften einer zeitlich modulierten turbulenten Kanalströmung*, Diploma thesis, University of Göttingen, Göttingen (1995).
[51] C. Völtz, *Zeitliche Modulation einer turbulenten Kanalströmung bei hohen Modulationsfrequenzen*, Diploma thesis, University of Göttingen, Göttingen (1997).
[52] P. Huerre and P. Monkewitz, 'Local and Global Instabilities in Spatially Developing Flows', Annu. Rev. Fluid Mech. **22**, 473 (1990).

Oscillations, Waves and Interactions, pp. 107–138
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-05-1

Active control of sound and vibration History – Fundamentals – State of the art

Dieter Guicking

Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Email: guicking@physik3.gwdg.de

Abstract. In a narrow sense, coherent active control of sound and vibration is the cancellation or (less often) enhancement by superposition of an antiphase or in-phase additional signal, usually from an external source of sound or vibration. The historical development of the technologies are outlined, the fundamentals under aspects of physics, signal processing and algorithms are treated, and the current states of research and applications are reviewed, more or less systematically sorted. Related fields such as adaptive optics, active flow control, and control of nonlinear dynamical systems are also included. Active control of sound and vibration in a wider sense, the incoherent superposition, aimed at sound power enhancement etc., is not considered in this survey.

1 Introduction

The concept of cancelling unwanted sound or vibrations by superimposing a compensation signal exactly in antiphase is not new. In acoustics, most of the early publications in this field are patent applications, showing that technical applications were considered possible. However, for a long time experiments were nothing more than laboratory demonstrations which were smiled at as curiosities, far from reality. Only modern electronics made technical applications feasible.

The situation was different with the compensation of low-frequency mechanical vibrations; these techniques were used in practise at a very early stage. In the following, an overview is given of the historical development, technical realisations, and present research activities.

2 Active noise control (ANC)

2.1 Early investigations

The first experiments on the superposition of sound fields were presumably made in 1878 by Lord Rayleigh [1]. He describes under the heading "Points of Silence" how he scanned, with his ear, the interference field produced by two electromagnetically

synchronised tuning forks, and that he found maxima and minima of loudness. Although it can be assumed that these experiments should only prove that coherent sound fields can interfere in the same way as do optical fields (which was known since the days of Thomas Young), patent applications by Coanda [2,3] and Lueg [4] were aimed at possible noise reduction – however only in Lueg's proposal in a physically realistic way. Lueg's German and especially the related US application [5] with an additional sheet of drawings are, therefore, considered rightly as the first written documents on active control of sound.

Lueg already proposed the usage of electroacoustic components, but the first laboratory experiments were documented by Olson 1953 [6] and 1956 [7], who also listed far-sighted prospective applications. Technical applications were not possible at that time because of the clumsy electronic vacuum tube equipment, lacking sufficient versatility. Also, our ears present a problem, namely the nearly logarithmic dependence of the perceived loudness on the sound pressure. For example, a sound level reduction by 20 dB requires an amplitude precision of the compensation signal within 1 dB and a phase precision within 6 degrees of the nominal values – for all frequency components of the noise signal. These demands, together with the requirement of temporal stability, have impeded for a long time the technical use of coherent-active compensation systems (also termed *anti-sound*) until in recent years digital adaptive filters proved to be the appropriate tool.

2.2 The energy objection

In the context of the active cancellation of sound fields a question often posed is "Where does the energy go?" With the seemingly convincing argument that the primary field energy can only be enhanced by adding secondary sound sources, the concept of active noise control (ANC) is principally questionable [8]. The objection is correct if the cancellation is achieved by interference only; a local cancellation leads to doubling of the sound pressure elsewhere. But a more detailed consideration reveals that the secondary sources can, properly placed and driven, absorb the primary energy. In other situations, the sources interact such that the radiation impedance is influenced and thereby the sound production reduced. This will be elucidated in the following sections.

2.3 The JMC theory

M. Jessel and his coworkers G. Mangiante and G. Canévet have developed a theory which has become known, after their initials, as the *JMC theory*. They have treated the problem sketched in Fig. 1 (e. g., Ref. [9]). Sound sources Q are located within a volume V with surface S. Along S, secondary sources shall be arranged such that they compensate the sound field radiated to the outside, but do not alter the field within V. This is possible according to Huygens' principle: substitute sources q continuously distributed along S can create the same sound field in the outside as the primary sources Q. With reversed poling, they produce a field which is in antiphase to the original one. Assuming that such reversed (and acoustically transparent) substitute sources operate together with Q, the sound fields in the outside cancel each other.

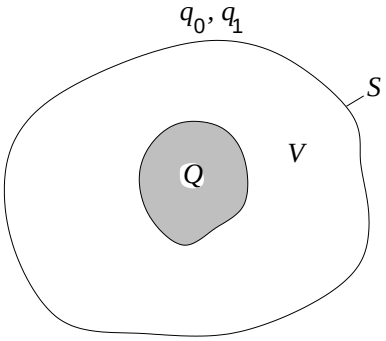


Figure 1. The Jessel–Mangiante–Canévet (JMC) theory.

If the cancellation sources are acoustic monopoles they radiate not only to the outside but also into V, creating standing waves and enhancing the sound energy in V. The inward radiation can be prevented by combining monopoles q_0 along S with dipoles q_1 so that the primary field in V is not altered. As to the energy, the *tripoles* formed by the q_0 and q_1 (directional radiators with cardioid characteristic) absorb, along S, the sound coming from Q. They serve as perfectly matched absorbers with an acoustic input impedance equal to the characteristic impedance of the medium.

With the same argument, it follows that a source-free region V can be shielded actively against sound influx from the outside by arranging appropriate compensation sources along the surface S of V. Monopole distributions along S reflect, tripoles absorb the incident sound.

For a given surface S and primary source distribution $Q(\boldsymbol{r})$, where $\boldsymbol{r}$ is the position vector, the substitution sources $q_0(\boldsymbol{r})$ and $q_1(\boldsymbol{r})$ can be calculated from the Helmholtz-Huygens integral equation which links the sound field in a region to the sound pressure and its gradient along the surface [10].

For practical applications, the theoretically required continuous source distribution has to be replaced by discrete sources. Their minimal surface density follows from their absorption cross section $A = \lambda^2/4\pi$ [11] and the smallest sound wavelength λ for which the system shall be effective. This concept has been verified in computer simulations [12] and experimentally in an anechoic room [13]. A practical application is noise shielding of large open-air power transformers by an array of loudspeakers to save the people living in the surroundings from the annoying hum [14]. It was also reported that cattle grazing near a noisy power transformer gave less milk.

A few researchers are further developing the JMC theory [15,16].

2.4 One-dimensional sound propagation and algorithms

Primary and cancellation sound must have the same direction of propagation. It is therefore easier to cancel plane, guided waves in ducts (below the cuton frequency of the first lateral mode) than, for example, three-dimensional sound fields in rooms with omnidirectional propagation. In a set-up as sketched in Fig. 2 (which was in

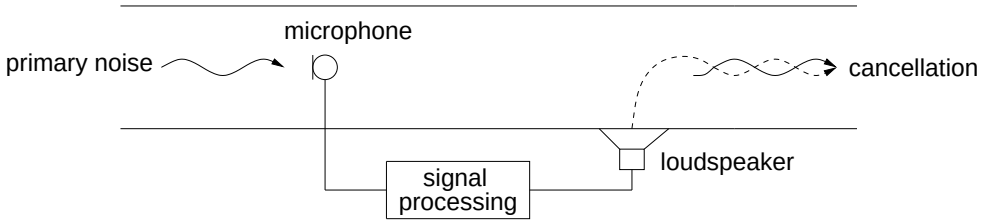


Figure 2. Principle of active feedforward cancellation of sound in a duct.

principle already proposed by P. Lueg [5] in 1934) the sound incident from the left is picked up by the microphone and, after some processing, fed to the loudspeaker such that to the right side the primary and the additional signal cancel each other.

After Lueg's idea, the "signal processing" should comprise the amplitude adjustment, sign reversal, and time delay according to the acoustic path length. However, an active noise control system is not practically applicable in this simple form. First, the acoustic feedback from the loudspeaker to the microphone has to be avoided and, second, in most cases it is necessary to follow up the transfer function adaptively since the time delay and the sound spectrum can change as a result of temperature drift, superimposed flow, and other environmental conditions. It is therefore common practise today to apply adaptive digital filters which are implemented on fast signal processors to enable online updating. Figure 3 shows a typical block diagram (amplifiers, A/D and D/A converters, and antialiasing lowpass filters being omitted).

The transfer function of the acoustic feedback path from the loudspeaker L to the reference microphone R is modeled by the feedback compensation filter FCF so that the input signal $x(t)$ to the main filter A does not contain contributions from L. The error microphone E receives, in the case of incomplete cancellation, an error signal $e(t)$ which serves for the adaptation of the filter A. This filter adapts such that it models the acoustic transfer function from R to L, including the (complex) frequency responses of R and L. The filters A and FCF are often realised as transversal filters (finite impulse response, or FIR filters), and the most common adaptation algorithm is the "filtered-x LMS algorithm" after Widrow and Hoff [17] where LMS stands for least mean squares. The algorithm is controlled by the product $e(t)x(t)$ and adjusts the filter coefficients by a stochastic gradient method so that $x(t)$ and $e(t)$ are decorrelated as far as possible. If the primary sound is broadband, the propagation delay from L to E decorrelates $x(t)$ and $e(t)$ to a certain degree which impairs the performance of the ANC system. In order to compensate for this effect, $x(t)$ is prefiltered in the update path (lower left) with a model $\tilde{H}_{LE}$ of the error path H_{LE}. The necessary error path identification is performed with an auxiliary broadband signal of the noise generator NG in the adaptation unit shown at the lower right of Fig. 3. The coefficients of $\tilde{H}_{LE}$ (and also of FCF) are either determined once at start-up and then kept constant or, if the transfer functions vary too much with time, permanently; in the latter case, however, the (weak) auxiliary signal remains audible at the duct end since it is not cancelled by the loudspeaker signal $y(t)$.

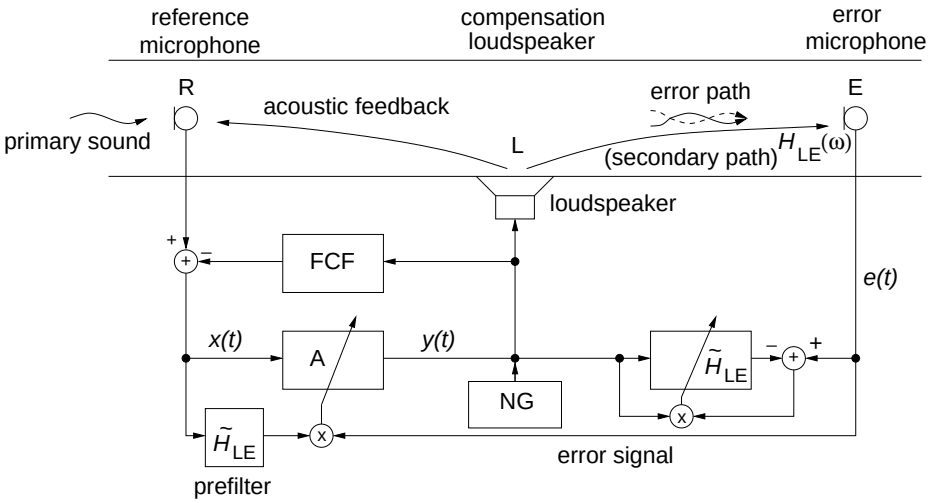


Figure 3. ANC in a duct by adaptive feedforward control with feedback cancellation and error path identification for the filtered-x LMS algorithm.

After adaptation, the loudspeaker acts as a sound-soft reflector for the wave incident from the left which is, hence, not absorbed but reflected to the left. With a different control strategy the loudspeaker can be operated as an "active absorber", but the maximum possible absorption is half of the incident sound power; either one quarter are reflected and transmitted. The reason is that it is not possible to achieve perfect impedance matching with a single loudspeaker mounted at the duct wall. The incident wave 'sees' the parallel connection of the loudspeaker input impedance and the characteristic impedance of the ongoing part of the duct. (But a loudspeaker at the end of a duct can be driven to perfectly absorb the incident sound [18].)

If the standing waves or the stronger sound propagation to the left in arrangements as those in Figs. 2 and 3 cannot be tolerated, a true active absorber can be realised with loudspeaker pairs or linear arrays [19,20].

A series of commercial ANC systems working on the principle of sound-soft reflection have been developed by the US company Digisonix and successfully installed mainly in industrial exhaust stacks since 1987 [21]. The filters A and FCF are combined to one recursive, infinite-impulse response (IIR) filter, often applying the Feintuch algorithm [22]. The signal processors allow on-line operation at least up to 500 Hz, suppress tonal noise by up to 40 dB and broadband noise typically by 15 dB. Similar systems have been installed also in Germany [23–25] and elsewhere [26]. The lower frequency limit is given by pressure fluctuations of the turbulent flow, the upper limit by the computational speed of the signal processor and the lateral dimensions of the duct. The higher modes occurring at higher frequencies can also be cancelled, requiring, however, a greater amount of hardware [27]; therefore, only few such systems with multi-mode cancellation have been installed so far.

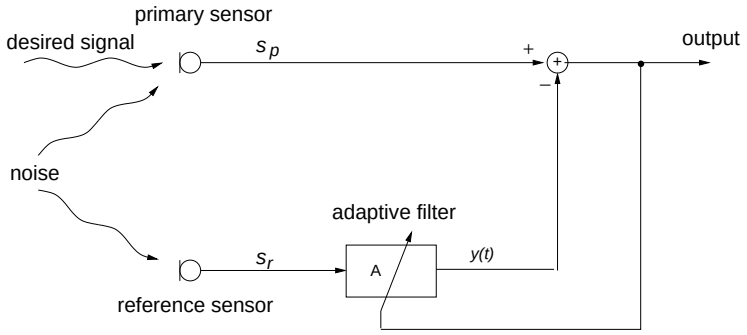


Figure 4. Adaptive noise cancelling.

The filtered-x LMS algorithm is very popular because of its moderate signal processing power requirement (the numerical complexity is $O(2N)$ if N is the filter length), but its convergence is very slow for spectrally coloured random noise. Fan noise spectra have typically a steep roll-off with increasing frequency so that the convergence behaviour of the algorithm is often insufficient. Efforts have therefore been made to develop algorithms the convergence behaviour of which is independent of the signal statistics, but which can still be updated in real time. One example is the SFAEST algorithm [28] which has a complexity of $O(8N)$. Since it furthermore calculates the optimal filter coefficients in one single cycle, it is particularly useful for nonstationary signals and nonstationary transfer functions. Stability problems in the initialisation period could be solved by the FASPIS configuration which stands for fast adaptive secondary path integration scheme [29,30]. More on algorithms can be found in the books [31] and [32]. The very difficult extension of the fast algorithms and the FASPIS configuration to IIR filters has been accomplished in the doctoral thesis of R. Schirmacher [33].

The modern control theory provides advanced algorithms such as H_∞, H_2, fuzzy control, optimal control, artificial neural networks, genetic algorithms, to name just a few. Overviews are presented, e. g., by the books [34] and [35].

An important concept in many fields of ANC is *adaptive noise cancelling* which became widely known since 1975 by B. Widrow et al.'s seminal paper [36], see Fig. 4. A 'primary' sensor picks up a desired signal which is corrupted by additive noise, its output being s_p. One or more 'reference' sensors are placed such that their output s_r is correlated (in some unknown way) with the primary noise, but does not essentially contain the desired signal. Then, s_r is adaptively filtered and subtracted from s_p to obtain a signal estimate with improved signal-to-noise ratio (SNR) since the adaptive filter decorrelates the output and s_r. This concept, realised by a linear predictive filter employing the least mean squares (LMS) algorithm, has been patented [37] and found wide applications: in speech transmission from a noisy environment [38], in seismic exploration [39], medical ECG diagnostics [40], noise cancelling stethoscopes [41], speech enhancement in noisy environment [42], hearing aids [43], and many other problems. In experimental cosmology, adaptive noise cancelling will be used in the

"Low-Frequency Array" (LOFAR) project for the detection of the 21-cm hydrogen radiation from the early universe, which – by cosmic expansion – has been shifted to about 2-m wavelength. The interference by the overwhelmingly stronger signals from terrestric radio transmitters shall be eliminated adaptively [44].

In adaptive feedforward control systems as shown in Fig. 3 the sound propagation path from microphone R to loudspeaker L must be long enough to provide the time required for calculating the signal to be fed to L (*causality condition*). The limiting factor is usually not the computation time in the signal processor but the group delay in the antialiasing lowpass filters which are necessary in digital signal processing.

Problems in technical ANC applications are often posed by the loudspeakers. Very high low-frequency noise levels are typically encountered in exhaust stacks or pipes, demanding for high membrane excursions without nonlinear distortion and, often, robustness against aggressive gases and high temperatures. On the other hand, a smooth frequency response function (as for Hi-Fi boxes) is not an issue because frequency irregularities can be accounted for by the adaptive filter. Special loudspeakers for ANC systems have been developed [45–48].

2.5 Interaction of primary and secondary sources

The ANC systems discussed in the preceding sections are aimed at absorption or at least reflection of the primary sound power, tacitly assuming that the primary power radiation is not influenced by the cancellation sources. However, if it is possible to reduce the primary sound production by the operation of the secondary sources, this will be a particularly effective method of noise reduction.

A monopole radiator of radius a with a surface particle velocity v produces a volume velocity $q_0 = 4\pi a^2 v$. The sound power radiated into a medium of density ρ and sound velocity c at a frequency ω (wavelength λ, wave number $k = 2\pi/\lambda$) is $P_0 = \rho\omega^2 q_0^2/(4\pi c)$. Adding an equal but antiphase monopole at a distance $d \ll \lambda$, produces a dipole which radiates the power $P_1 = P_0(kd)^2/3$. Supplementing this dipole with another one to form a quadripole, the radiated power is further reduced to $P_2 = P_0(kd)^4/15$, assuming $kd \ll 1$ [49, Chapter 7.1].

These conditions are correct if the volume velocity q is the same in all three cases, but this is not necessarily so because ANC, by adding a compensation source in close proximity, does not only raise the multipole order, but can also alter the radiation impedance $Z_r = R_r + j\omega M_r$. The surrounding medium acts upon a monopole with the radiation resistance $R_{r0} = 4\pi a^2 \rho c$ and the mass load $M_{r0} = 4\pi a^3 \rho$ (three times the replaced fluid mass), upon a dipole with $R_{r1} = R_{r0} \cdot (ka)^2/6$ and $M_{r1} = M_{r0}/6$, and on a quadripole with $R_{r2} = R_{r0} \cdot (ka)^4/45$ and $M_{r2} = M_{r0}/45$ [49]. The mass load leads to a reactive power, an oscillation of kinetic energy between primary and secondary source ("acoustical short-circuit"). The product $v^2 R_s$ determines the radiated (active) power. The particle velocity v of the primary source depends on its source impedance and the radiation resistance. A "low-impedance" source (sound pressure nearly load independent) reacts on a reduced radiation resistance R_s with enhanced particle velocity v so that the reduction of radiated power by the higher multipole order is partially counteracted. But the sound radiation of impedance-matched and of "high-impedance" (velocity) sources is reduced in the expected way

by an antiphase source in the nearfield.

These relationships can be utilised, e. g., for the active reduction of noise from exhaust pipes (ships, industrial plants, automobiles with internal combustion engines). A large demonstration project was implemented as early as 1980: the low-frequency hum (20 to 50 Hz) from a gas turbine chimney stack was cancelled actively by a ring of antisound sources [50]. Each loudspeaker was fed from one microphone pair through amplifiers with fixed gain and phase settings. Such a simple open-loop control was sufficient in this case due to the highly stationary noise and its narrow frequency band.

The pulsating gas flow emanating from a narrow exhaust pipe is a very efficient "high-impedance" monopole sound radiator; an adjacent antiphase source turns it into a dipole or, in the case of a concentric annular gap around the exhaust mouth, into a rotationally symmetric quadripole. Such "active mufflers" for cars have often been proposed (e. g., Ref. [51]), but practical installations are still lacking, for technical and economical reasons: microphones and loudspeakers beneath the car body must be protected against shock and vibration, splash water, thrown-up gravel and the hot, aggressive exhaust gas [52]. Furthermore, active mufflers have to compete with the highly efficient and comparatively cheap conventional mufflers from sheet metal. Researchers in the muffler industry are, however, still developing and improving active systems, testing prototypes, and they are optimistic that active mufflers might go into production because they combine noise cancellation with backpressure reduction; perspectives to include sound quality design are seen, too [53].

2.6 Waveform synthesis for (quasi)periodic noise

A conceptually simple adaptive algorithm has been developed by a British research team [51]. It assumes (quasi)periodic noise, the source of which is accessible for obtaining synchronisation pulses (e. g., vehicle engine noise). The principle is explained in Fig. 5. A loudspeaker is mounted next to the exhaust pipe end, and is fed from a *waveform synthesiser*, realised with digital electronics. An error microphone is placed in the superposition zone and yields a control signal by which the loudspeaker output is optimised. The sync pulses (obtained, e. g., by a toothed wheel and an inductive probe) guarantee that the compensation signal tracks the changing engine rotation speed automatically. The waveform is adapted using a trial-and-error strategy either in the time domain or, faster, in the frequency domain. In the latter case, the amplitudes and phases of the (low order) harmonics of the engine noise are adapted. The prominent feature of this active system is that no microphone is required to receive the primary noise because the signal processor performs the waveform synthesis by itself. The loudspeaker must only provide the necessary acoustic power; resonances, nonlinearities and ageing are automatically accounted for. A disadvantage is the slower convergence as compared to "true" adaptive algorithms.

An example for a technical application of ANC with waveform synthesis in medicine is a noise canceller for patients undergoing a magnetic resonance imaging (MRI) inspection. The electrical high-current impulses that are needed to build up the required high magnetic fields cause, by magnetostriction and "wire forces", an annoying impulsive noise which is cancelled with the help of an active headset (see Section 2.7).

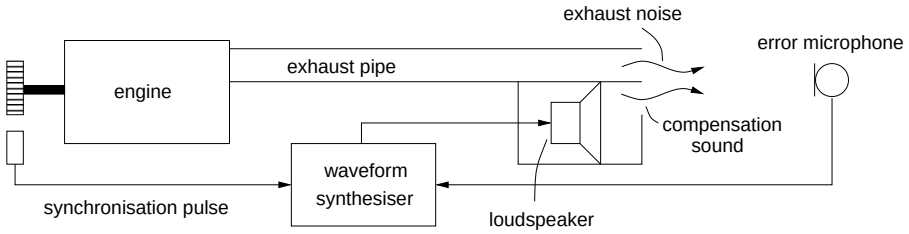


Figure 5. Active cancellation of (quasi)periodic noise by tracking control with sync input and waveform synthesis (after Ref. [51]).

Because no ferromagnetics and preferably no metal at all must be brought into the MRI tube, pneumatic headsets with long plastic tubes as sound guides have been developed for this purpose which are fed from a signal processor with a simple feedforward control and fixed filters [54]. Since, however, the compensation is not very good, an improvement with a metal-free optical microphone for controlling an adaptive filter has been developed [55–57]. A different approach is aimed at controlling the structural vibrations of the MRI tube walls [58,59].

2.7 Small volumes – Personal noise protection

An acoustically simple ANC problem is presented by an enclosure the dimensions of which are small compared with the wavelength even at the highest frequencies of interest. The sound pressure is then spatially almost constant, and the cancelling source can be placed anywhere in the enclosure. Correctly fed, it acts as an active absorber.

One such small enclosure is the space between a headphone and the ear drum. The concept of "personal noise protection" by actively controlled headphones was originally claimed in a Russian patent application [60], but reliable signal processing was, in spite of intense research work in many countries, possible only very much later. Independent developments by the US company Bose [61] and Sennheiser in Germany [62] resulted in active headsets for aircraft pilots; active headsets are meanwhile also produced by other companies, being offered as pure hearing protectors in open or closed construction, with feedforward and feedback control in analog electronics, and also with a signal input for telecommunication. The initially very costly active headsets have become so much cheaper that a wider application in vehicles and noisy working places appears realistic. Quite recently, also adaptive digital signal processing has been applied to active headsets and hearing protectors [63].

2.8 Local cancellation

Placing an anti-source in the immediate nearfield of a primary noise source gives a "global" effect as explained in Section 2.5, but if the distance of the two sources gets wider, then only a local cancellation by interference remains [11]. Such systems did not receive general attention as noise cancelers because of their very limited spatial

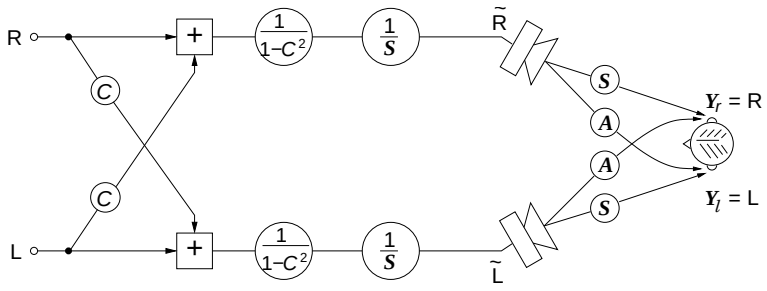


Figure 6. Crosstalk cancellation in head-related stereophonic sound field reproduction with two loudspeakers by prefiltering. $S = S(\omega)$ and $A = A(\omega)$ are the nearside and farside transfer functions, respectively, between loudspeakers and eardrums; circles to the left from the loudspeakers indicate filters with inscribed transfer functions, $C = C(\omega) = -A(\omega)/S(\omega)$.

range of efficiency (in the order of $\lambda/10$). But local cancellation can be very useful for acoustic laboratory experiments, such as head-related stereophony when dummy head recordings are reproduced by two loudspeakers [64]. As the sound radiated from the left loudspeaker should be received by the left ear only, a compensation signal is superimposed onto the right channel which compensates the sound coming from the left loudspeaker to the right ear, and vice versa, see Fig. 6. As compared to the familiar source localisation between the loudspeakers of a conventional stereo set, this procedure provides true three-dimensional sound field reproduction with source localisation in any direction, including elevation, and also gives a reliable depth impression.

Of great practical relevance is local active sound field cancellation for teleconferencing and hands-free telephones (speakerphones) in order to eliminate, at the microphone location, acoustic room echoes which degrade the speech quality and tend to cause howling by self-excitation; the active system causes dereverberation of the room response [65,66]. Echo cancellation and a speech enhancement system for in-car communication are described in Ref. [67]. Echo cancellation for stereophonic sound field reproduction is more involved than single channel applications. Solutions are presented, e. g., in Refs. [68] and [69]. The psychoacoustic aspect of masking was introduced in acoustic echo cancellation combined with perceptual noise reduction [70]. For echo cancellation in fast changing environments, a special algorithm has been developed [71].

A hot topic in speech transmission with multiple not precisely known sound sources is *blind source separation*, using microphone arrays and algorithms such as spatial gradient estimation, independent component analysis (ICA), statistical source discrimination, maximum likelihood, and Kalman filters; Ref. [72] presents a comprehensive survey.

A related older problem is the removal of electric line echoes in long-distance telephony with satellite communication links where the long transmission path leads to audible echoes which greatly disturb speech communication [73]. The signals are reflected from an impedance mismatch at the so-called hybrid where the two-wire line

branches into the four-wire local subscriber cable. The geostationary satellites are positioned at 36 000 km height so that the echo return path (transmitter → satellite → receiver → satellite → transmitter) is 4×36 000 km which yields, in spite of the signal propagation at the speed of light, an echo delay time of as much as nearly 0.5 s. All satellite telephone links are therefore equipped with transmission line echo compensators (see, e. g., Ref. [74]).

Locally effective ANC systems with compact microphone/loudspeaker systems in feedback configuration were described by H. F. Olson [7] as early as 1956; they absorb low-frequency sound in a narrow space around the microphone and were proposed for aircraft passengers and machine workers [75]. Because of the very restricted spatial field of efficiency, such systems did not receive general attention. In more recent experiments the test persons disliked also the strong sound level fluctuations when they moved their head.

The application of acoustic echo cancellation was also proposed for ultrasonic testing where flaw echoes can be masked by strong surface echoes. It is possible to subtract the latter from the received signal and so improve the detectability of flaws [76,77]. Similarly, the ANC technique can be applied to cancel the reflection of the ultrasonic echo from the receiver [78].

2.9 Three-dimensional sound fields in enclosures

The active cancellation of complex sound fields in large rooms, possibly with non-stationary sources and time-varying boundary conditions, is far beyond the scope of present ANC technology. More realistic is the concept of reducing room reverberation by placing active absorbers along the walls. The incident sound is picked up by microphones which feed the loudspeakers so that their acoustic input impedance is matched to the sound field. The situation is the same as in Fig. 1 if the enclosure walls are considered as a Huygens surface. The loudspeakers can also be driven such that their reflectivity takes arbitrary values in a wide frequency range (experimentally, reflection coefficients between 0.1 and 3 have been realised). This would facilitate the construction of a room with adjustable reverberation time [79], but at present still with a prohibitive amount of hardware.

The concept of active impedance control was originally propagated by our Göttingen team [80,81] and has stimulated many later research activities, (e. g., Refs. [82–86]).

Intensive research has been devoted to the active cancellation of sound in small enclosures such as vehicle, aircraft and helicopter cabins. Four-stroke internal combustion engines have an inherent unbalance at twice the rotational speed (the "second engine order") which often coincides with the fundamental cabin resonance of cars, so exciting the highly annoying "boom". Since this noise is strongly synchronised with the engine speed its active cancellation is possible with a relatively small amount of hardware and software [87]. It was, however, only offered in a production car for some time by NISSAN for their model *Bluebird* in Japan. Many other car manufacturers develop their own systems, and some of them have successfully built prototypes, but all of them are hesitating to install the ANC systems in series production (e. g., within a "comfort package" at extra cost). One argument is that customers would complain if they pay for noise reduction, and there still remains some disturbing noise.

More involved than the "boom" control is the cancellation of the broadband rolling noise, both inside and outside the car. Laboratory experiments and driving tests have led to preliminary solutions; the nonstationarity of the noise input and of the acoustic transfer functions demand for fast adapting algorithms, also for the error path identification [29,33,88,89]. The noise and vibration problems are becoming more severe with small low-consumption cars now under development; they will possibly be equipped with both active noise control for the interior space and active vibration control for the engine and wheel suspensions. For more luxurios cars the trend in the automobile industry goes to combining ANC technology with "sound quality design" for the car interior so that the driver has the choice, e. g., of a more silent car or a more sportive sound [90–93].

For economical reasons, the aircraft industry has replaced jet engines by propeller (or turboprop) aircraft for short and medium distances which are, however, much louder in the cabin. Relatively little effort is necessary to employ a technology known as *synchrophasing*. The eddy strings separating from the propeller blade tips hit the fuselage and excite flexural vibrations of the hull which radiate sound into the cabin. If the right and left propeller are synchronised so that their "hits" meet the fuselage out of phase instead of simultaneously, then higher-order shell vibrations are excited which radiate less and so reduce the noise level inside [94]. Better results, however with more involved installations, are obtained with multichannel adaptive systems. An international European research project with the acronym ASANCA has resulted in a technical application [95].

An important issue in ANC applications to three-dimensional sound fields is the placement of microphones and loudspeakers. Attention has to be paid not only to causality, but also to observability and controllability, in particular in rooms with distinct resonances and standing waves (*modal control*). If, for some frequency, the error microphone of an adaptive system is positioned in a sound pressure node, it does not receive the respective frequency component or room mode so that no cancelling signal will be generated and no adaptation is possible. If the loudspeaker is placed in a node, then a compensation signal calculated by the processor cannot be radiated effectively into the room, which usually forces the adaptive processor to produce higher and higher signal amplitudes, finally leading to an overload error of the digital electronics.

2.10 Freefield active noise control

Technical applications of ANC to three-dimensional exterior noise problems are still quite rare, but many research projects have been reported and a number of patents exist. The problems with active mufflers for cars with internal combustion engines have been discussed in Section 2.5. A technically similar problem is the fly-over noise of propeller aircraft which mainly consists of two components: the propeller blade tip vortex threads, and the equally impulsive exhaust noise. If the exhaust tail pipe is shifted to a position near to the propeller plane, and if the angular position of the propeller on its shaft is adjusted so that in downward direction the pressure nodes of one source coincide with the antinodes of the other one, then the destructive interference reduces the fly-over noise by several dB [96].

A method for reducing traffic noise by cancelling the tyre vibrations of an automobile is disclosed in a patent [97], proposing electromagnetic actuation of the steel reinforcement embedded into the tires.

A frequently investigated problem is the cancellation of power transformer noise, the annoying hum of which consists of multiples of the power line frequency (50 Hz, in USA 60 Hz). It is a seemingly simple problem because of the strong periodicity and the readily accessible reference signal. Several methods have been proposed, either by loudspeakers arranged around the site [98], by force input to the oil in which the transformer is immersed [99] or to the surrounding tank walls [100], or by sound insulating active panels enclosing the transformer [101]. Experimental results are discussed in Ref. [102]. Problems are posed, however, first, by the weather-dependent sound propagation – wind and temperature gradients tilt the wave front [103] – and second, because the hum spectrum depends on the electrical load of the transformer [104].

It has also been tried to actively improve sound shielding noise barriers along roads, in particular to cancel the low frequency noise diffracted around the barrier top. The idea is to place loudspeakers along the upper edge and to drive them with adaptive feedforward control, the reference microphones being placed on the roadside and the error microphones in the shadow zone [105]. Improvements are concerned with multiple loudspeaker arrays also along the side walls of the noise barrier [106], or multiple reference control and virtual error microphones [107].

3 Active vibration control (AVC)

3.1 Early applications

In contrast to active noise control, active vibration control has long been applied, in particular to ships. Ref. [108] (1905) reports on vibration reduction on a steam ship by synchronisation of the two engines in opposite phase, Ref. [109] (1934) on the reduction of roll motion by an actively driven Frahm tank (water is pumped between tanks located on the two sides of the ship), and Ref. [110] (1945) on roll stabilisation by buoyancy control with "activated fins", auxiliary rudders with variable angle of attack protruding laterally from the ship hull into the water. The latter technology is still applied today.

Active damping of aircraft skin vibrations was proposed by Ref. [111], providing multichannel feedback control with displacement sensors and electromagnetic actuators, mainly in order to prevent fatigue damage.

Early publications can also be found on the active control of vibrations in beams, plates and composite structures. In mechanical wave filters where a desired longitudinal wave mode in a bar is superimposed by an interfering detrimental flexural wave mode, the latter can be damped by pairs of piezoelectric patches on either side of the bar which are connected through an electrical resistor [112].

In special environments, e. g. ultrahigh vacuum, magnetic bearings without lubricants are preferred for rotating machinery, but their inherent instability requires feedback control which equally reduces vibrations [113].

An early NASA patent [114] provides an active mass damper (see Section 3.4) to cancel structural vibration.

In the 1980s, longitudinal vibrations of the ship superstructure caused by nonuniform propulsion were reduced with a type of dynamic absorber, realised by a *centrifugal pendulum*. This is a pendulum swinging along the length direction of the ship and rotating about an axis pointing also in lengthwise direction. The swinging of the pendulum is synchronised to the ship's vibration by controlling the rotational frequency, and hence the centrifugal force, which together with gravity determines its natural frequency [115].

3.2 AVC for beams, plates and structures

Aircraft and spacecraft have a great impact on investigations in active control of structural vibration. Other than the above-mentioned whole-body vibrations of ships which are comparatively easy to control due to their very low frequencies, one is here confronted with elastic structures, i. e., continuous media with an infinite number of degrees of freedom the control of which presents fundamentally different problems. First, there are the different wave types in solids (of which longitudinal, torsional and transversal waves are the most important); their control demands for various types of actuators and sensors. Furthermore, the propagation speed is generally higher in solids than in air so that causality problems occur with broadband adaptive feedforward controllers. As a consequence, many problems are treated with modal control where, especially in case of overlapping modes, the "control spillover" problem has to be considered: the unwanted excitation of additional modes the resonance curves of which extend to the controlled frequency. In Fig. 7 it is assumed that the Nth mode, resonant at frequency f_N, shall be controlled; the tails of the neighbouring resonance curves have nonnegligible amplitudes at f_N (the dots on the dashed line) and are therefore also excited by the control signal at f_N, to some extent. Owing to the phase slope around a resonance, the neighbouring modes are usually enhanced rather than damped when the Nth mode is suppressed. While control spillover leaves the system stable, "observation spillover" can produce instability [116].

For satellites, the damping of modal vibrations is important after pointing position manoeuvres etc. since they are built from low-loss materials, and air friction is not present in space. The optimisation of number and placement of sensors and actuators for the mostly applied adaptive feedback controllers requires precise knowledge of the structural dynamics so that reliable modelling in state-space coordinates and a realistic estimation of discretisation errors are possible. An introduction to this field is given by Meirovitch [116].

Damping and stiffness control in mechanical junctions can also be achieved by dry friction control where the pressing force is controlled by a piezoelectric actuator, in feedforward or feedback control, typically by a nonlinear algorithm, e. g. a neural network [117], or by on–off control [118].

In aircraft technology, active controllers have been developed for manoeuvre [119] and gust load alleviation [120], as well as for wing flutter control [121], and for noise and vibration reduction in helicopters [122], in particular by individual blade control (IBC) [123] and higher harmonic control (HHC) [124]. A major German research initiative was the "Adaptive Wing Project" [125], aimed at aircraft drag reduction by boundary layer and flow separation control with the help of wing shape control,

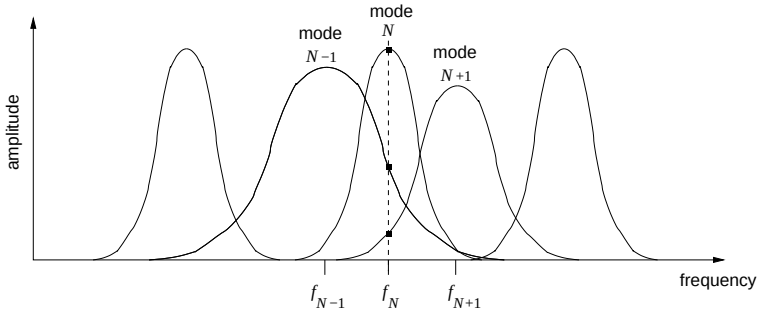


Figure 7. Spillover: attempting to control mode N at its resonance frequency f_N also excites neighbouring modes $N-1$ and $N+1$ with lower, but finite amplitudes.

realised by an adjustable lengthwise tiny bump near the trailing edge of the wing, with piezoelectric or shape memory alloy actuators.

Initially, technical problems were encountered, among others, by the fact that sensor and actuator materials such as piezoceramics, piezopolymers, electro- and magnetostrictive materials, shape memory alloys, electro- and magnetorheological fluids are no constructional materials with a mechanical strength sufficient for load-bearing structures; some of them are also too brittle or too weak for fail-safe operation. This led to a new research field since the end of the 1980s: the development of modern compound materials with embedded sensors and actuators (keywords are intelligent (or smart) materials, bi-functional elements, adaptive (or smart) structures, adaptronics, structronics) [126–128]. Much information on these research fields is published in the special journals "Journal of Intelligent Material Systems and Structures" (since 1990), "Smart Structures and Materials" (since 1992), in the Proceedings of the "International Conferences on Adaptive Structures and Technologies (ICAST)", and of the "Adaptronic Congresses" held in Germany since 1996.

Active vibration control has found a popular application in digital cameras with *image stabilisation*. The image blur by camera shake during the exposure is avoided by actively shifting the position of the CCD chip with a piezo-actuator, in response to a motion sensor signal [129].

3.3 Active vibration isolation

Possibilities for active noise control in road vehicles have been discussed in Section 2.9. The predominant sources of interior noise are engine and wheel vibrations which propagate as structure-borne sound through the car body and finally radiate airborne sound into the cabin. It is therefore reasonable to develop active engine mounts and active shock absorbers which are stiff enough to carry the static load, but dynamically resilient so that vibrations are not transmitted. Piezoceramic actuators are suited for excursions in the submillimetre range [130]; for larger amplitudes and forces at frequencies of a few Hertz, hydraulic and pneumatic actuators are available [131]. Compact and robust combinations of conventional rubber mounts

with electrodynamically driven hydraulics have been constructed as "active hydromounts" for a wide frequency range [132]. Active mounts are, for example, standard components of the DaimlerChrysler Mercedes CL Coupé [133].

In helicopter cabins, the principal noise source is the gear box, the vibrations of which are transmitted through typically 7 struts to the cabin roof (as structure-borne sound), and then radiated into the cabin as airborne sound. Particularly annoying are tonal components between 700 Hz and 4 kHz. The vibration transmission has been reduced by piezoelectric actuators at the struts so that the noise level in the cabin became much lower, as was verified in ground tests. The development towards a technical product is a current research topic [134].

Active control technology was applied for improved vibration isolation of tables for optical experiments, scanning microscopes, vibration sensitive semiconductor manufacturing stages, etc. Commercial products are offered by several companies, e. g., Newport (USA), Technical Manufacturing Corporation (TMC, USA), Halcyonics (Germany), and Integrated Dynamics Engineering (IDE, Germany); the latter company also offers active compensation systems for magnetic strayfields which is important for high resolution electron microscopes etc. Information is available from the companies' homepages.

For satellite missions, sophisticated controllers have been designed to actively isolate facilities for microgravity experiments from structural vibrations which are caused by position controllers and other on-board machinery [130,135].

The performance of hydraulic shock absorbers can be improved by applying electrorheological fluids (ERF) [136]. ERF are fine suspensions of polarisable small dielectric particles in an unpolar basic fluid, e. g., polyurethane in low-viscosity silicone oil [137]. Their viscosity can be adjusted reversibly between watery and pasty by applying electrical fields of several kV/mm.

Also suitable are magnetorheological fluids (MRF), suspensions of small ferromagnetic particles in a basic fluid, requiring a magnetic field for the viscosity to be changed. The field is usually applied by electromagnets which require a high electric current instead of a high voltage [138]. In order to provide a wide range of viscosity control, the viscosities of the basic fluids selected for ERF and MRF are as low as possible, which leads to sedimentation problems, in particular with MRF because of its specifically heavier particles than in ERF.

Nevertheless, much research is focusing on MRF applications: earthquake protection of buildings [139], journal bearings of rotating machinery [140], truss structures in spacecraft [141], vehicle suspensions [142], sandwich beams [143], cable swaying [144], adjustable dynamic absorbers for flexible structures [145], squeeze film dampers [146], and many other systems.

3.4 Civil engineering structures

Wind-induced swaying of tall, high-rise buildings can amount to amplitudes of several metres in the upper floors. This low-frequency sway can be reduced by tuned mass dampers (TMD) acting as resonance absorbers: masses of about 1% of the total mass of the building are placed on the top floor and coupled to the building structure through springs and dampers. Their performance is raised by actively enhancing the

relative motion. A prominent example where such an *active TMD* has been installed is the Citycorp Center in New York [147]. Less additional mass is required for aerodynamic appendages, protruding flaps that can be swivelled and utilise wind forces like sails to exert cancelling forces on the building [148].

Many research activities in the USA, Canada and in particular Japan are aimed at the development of active earthquake protection for buildings where, however, severe technical problems have still to be solved [149,150].

For slim structures such as antenna masts, bridges etc., tendon control systems have been constructed for the suppression of vibrations by controlled tensile forces acting in different diagonal directions [151,152].

3.5 Active and adaptive optics

The quality of pictures taken with optical or radio astronomical mirror telescopes depends essentially on the precision to which the optimal mirror shape is maintained. Modern swivelling large telescopes suffer from deformation under their own weight which is compensated more efficiently by active shape control than by additional stiffeners which inevitably enhance the mass of the structure. This technology is called *active optics* [153,154].

While the telescope motions are very slow (time constants above 0.1 s) and therefore easy to control, *adaptive optics* have solved the more complicated problem of controlling picture blurring by atmospheric turbulence, the so-called *seeing* which fluctuates at frequencies about 1000 Hz. The large primary mirror is fixed, but the smaller secondary mirror surface rests on a matrix of piezoceramic actuators which are adjusted by an adaptive multichannel controller so that a reference star is optimally focused. If no reference star exists in the vicinity of the observed object an *artificial guide star* can be created by resonance scattering of an intense laser beam from sodium atoms at about 100 km height [155,156]. Adaptive optics have improved the optical resolution of the best infrared telescopes by a factor of 10 to 50, to almost the diffraction limit.

This technology was developed in the USA during the 1970s for the military SDI project and has been declassified not before 1991 when civil research had reached almost the same state [157]. Meanwhile, this technology is applied to nearly all modern large optical infrared telescopes such as the Gemini North Telescope on top of the Mauna Kea on Hawaii [158] and the Very Large Telescope (VLT) in Chile, and will be applied to even larger telescopes planned for the future [159,160].

Adaptive optical mirrors have also found applications in industrial production for laser cutting and welding [161], and generally for optimising the quality of high-intensity laser beams [162,163]. Other non-astronomical fields of adaptive optics application are confocal microscopy [164], spatial light modulators (SLM) for optical telecommunication [165], and ophthalmology [166]. Most of the small deformable mirrors are manufactured as micromechanical systems (MEMS) (e. g., Ref. [167]). A survey of industrial and medical applications of adaptive optics is presented in Ref. [168]. The growing importance of this field can also be seen in the fact that many textbooks on adaptive optics have been published [169–173].

3.6 Noise reduction by active structural control

Active control of structural vibrations and active control of sound fields have been developed almost independently, including differing control concepts (mostly feedforward in acoustics, mostly feedback in vibration). But since some time the two fields have become connected. Many noise problems result from radiation of structure-borne sound, e. g. into the interior of cars and aircraft, on ships, by vibrating cladding panels of machines, etc. Here comes into action a concept known under the acronym ASAC (Active Structural Acoustic Control) [174] where noise reduction is not attained by superimposing airborne sound to the disturbing noise field but by controlling the vibrating structure itself. This is possible by suitably placed and controlled actuators to suppress the structural vibration, although this is not necessarily the optimal solution.

Acoustically relevant are mainly plate bending waves which due to their frequency dispersion ($c_B \propto \sqrt{\omega}$) are non-radiating at low frequencies and strongly radiating above the *critical frequency* ω_g at which the bending wave velocity equals the sound velocity c_0 in the surrounding medium. If $c_B < c_0$, the acoustical short-circuit between adjacent wave crests and troughs yields a weak sound radiation into the far-field, but for $c_B > c_0$ a very effective radiation results. The proportionality factor in $c_B \propto \sqrt{\omega}$ contains the flexural stiffness so that its modification shifts the critical frequency and can turn radiating modes into non-radiating ones (*modal restructuring*). Much work has been done to investigate how, e. g. by laminates from sheet metal and piezolayers as sensors and actuators, adaptive structures can be constructed which can suppress, in propeller aircraft etc., the above-mentioned fuselage excitation by eddy threads, so enabling a substitution for or at least a supplement to the more involved (and heavier) direct noise control by mirophone/loudspeaker systems [175–177].

3.7 Sound transmission control

Sound transmission through walls, windows, sound shielding plates etc. is effectively controlled by active means. This is often achieved by ASAC (see preceding section), but in some instances also by different means. Experiments have shown that sound transmission through double-glazed windows can be reduced by actively controlled loudspeakers in the gap between the glass panes [178]. Actively controlled double wall partitions are also reported in Refs. [179] and [180], the latter one for insulating floor impulsive noise.

The favourite actuators for active structural damping are piezoceramics, bonded to the structure to form adaptive (smart) structures [181]. Semi-active approaches apply passive (sometimes actively controlled) shunts across the piezoactuators to save energy [182,183], or even to gain electrical energy from the vibrated piezos, a rather new technology labelled “energy harvesting” or “energy scavenging” [184,185].

3.8 Control of nonlinear dynamical systems

The control of nonlinear dynamical systems has gained much attention in recent years, due to the great potential of applications in physics, enginering, medicine, and communication.

Of practical importance is the control of magnetic bearings to stabilise a rotor in its unstable equilibrium by feedback control. Being frictionless and free from lubricants, magnetic bearings are often applied in vacuum apparatus (also in spacecraft) such as high-speed centrifuges (e. g., Ref. [186]).

A particular realm of research is *chaos control*, forcing a chaotic oscillation into a stable periodic orbit [187,188]. Major control concepts are 1) feedforward control [189]; 2) feedback control by applying small perturbations to an accessible system parameter when the trajectory comes close to the unstable periodic orbit where it is desired to stabilise the system, the so-called OGY control, named after the protagonists of this method [190]; 3) *Time Delay Autosynchronisation* or *Delayed Feedback Control*, where the feedback signal is the difference of the actual and a previous output signal of the chaotic system [191]; and 4) *sliding mode control* [192]. An early form of the delayed feedback control concept was formulated in the theory of balancing rods by humans and bicycle riding [193].

A medical application is the stabilisation of atrial fibrillation, a chaotic rapid oscillation of blood flow in the heart vestibules [194].

A potential application of chaos control is secure communication by masking the message with a broadband chaotic carrier at the transmitter site and demasking it at the receiver site by synchronising the chaotic transmitter and receiver oscillators [195].

An extension of delayed feedback control is *Multiple Delay Feedback Control* where the feedback signal contains more than one previous observables. This concept was applied successfully, e. g., to the stabilisation of a Colpitts oscillator and a frequency-doubled solid state laser [196].

Related to chaos control is *bifurcation control*; an overview outlining the theory, control concepts, and potential applications is given in Ref. [197].

While most applications of chaos control are aimed at converting an unpredictable process to a regular one, some other situations favour the transition of regular to chaotic behaviour (*anticontrol* or *chaotification*), such as in combustion engines where chaos (here: turbulence) enhances the mixing of fuel and air and so leads to better performance. Another example where the forced transition of regular oscillation into chaotic motion is beneficial is the improvement of a neural network by state feedback control [198].

For more information about the subject see the article of U. Parlitz in this book [199].

4 Active flow control

Coherent active control technology is also applied to other fields than sound and structural vibrations, among which the physics of fluid flow is gaining more and more importance. One of the many interactions of sound and flow is the transition from laminar flow of a slim gas flame into turbulence by insonification. Conversely,

the turbulence of a flame was suppressed actively by feedback control with a microphone/loudspeaker system [200].

Laboratory experiments have shown since about 1982 that the transition from laminar to turbulent flow can be shifted to higher Reynolds numbers by controlling the Tollmien-Schlichting waves in the boundary layer, thereby providing drag reduction which is of great technical relevance. This can be achieved with thermal inputs [201], or by acoustical or vibrational excitation [202–204].

Also, the dangerous surge and stall in compressors, resulting from instabilities, can be suppressed acoustically [205].

An ionised gas stream in a combustion chamber (as in a rocket) tends to produce unstable resonance oscillations which can be suppressed by an appropriately controlled electric d.c. current through the ionised gas, employing a feedback controller with a photoelectric cell as oscillation sensor [206].

A micro-electromechanical system (MEMS) to be mounted on fan blades is presented in Ref. [207], comprising a turbulence sensor, an integrated circuit, and an actuator by which turbulence noise can either be reduced, or – in the case of heat exchangers – amplified in order to improve heat transfer. Experiments on this interesting technique are described in Ref. [208], reporting flight control of a delta wing aircraft, and in Ref. [209] where the laminar/turbulent transition is influenced along the wing profile in a wind tunnel.

Blade-vortex interaction causing the rattling impulsive noise from helicopters can be reduced by controlling flaps at the trailing edges [210]. Also, helicopter stall can be controlled by trailing-edge flaps [211], or by plasma actuators [212]. In a further development, tip vortices of helicopter blades, aircraft foils, or marine propellers can be reduced by air injection to the high-pressure side of the lifting body [213,214].

Dynamic stabilisation of jet-edge flow with various adaptive linear feedback control strategies was experimentally verified by Ref. [215]. Disturbing resonances in a large wind tunnel with free-jet test section (the so-called Göttingen model) can be suppressed by feedback ANC, employing multiple loudspeakers [216]. Active flow control can also provide a low-frequency high-intensity sound source, utilising an aeroacoustic instability [217,218].

Conclusions

Coherent active control systems are commercially applied in acoustics in certain problem areas, but only in acoustically somehow “simple” situations: small volume, one-dimensional sound propagation, quasiperiodic noise, isolated modes. There are many more applications in vibration technology, but there are also fields where the non-application of a well-developed technique is at first sight surprising, among them flutter control of aircraft wings which has been successfully tried since more than 30 years. Here the flutter limit would be shifted to a higher flight speed, but this is not practised from safety considerations: if the active controller fails, and that cannot be excluded with such complex systems, the danger of wing fracture and hence an air crash would be too high. This is a general problem; precautions have to be taken in safety-relevant applications where failure of the active control system must not have

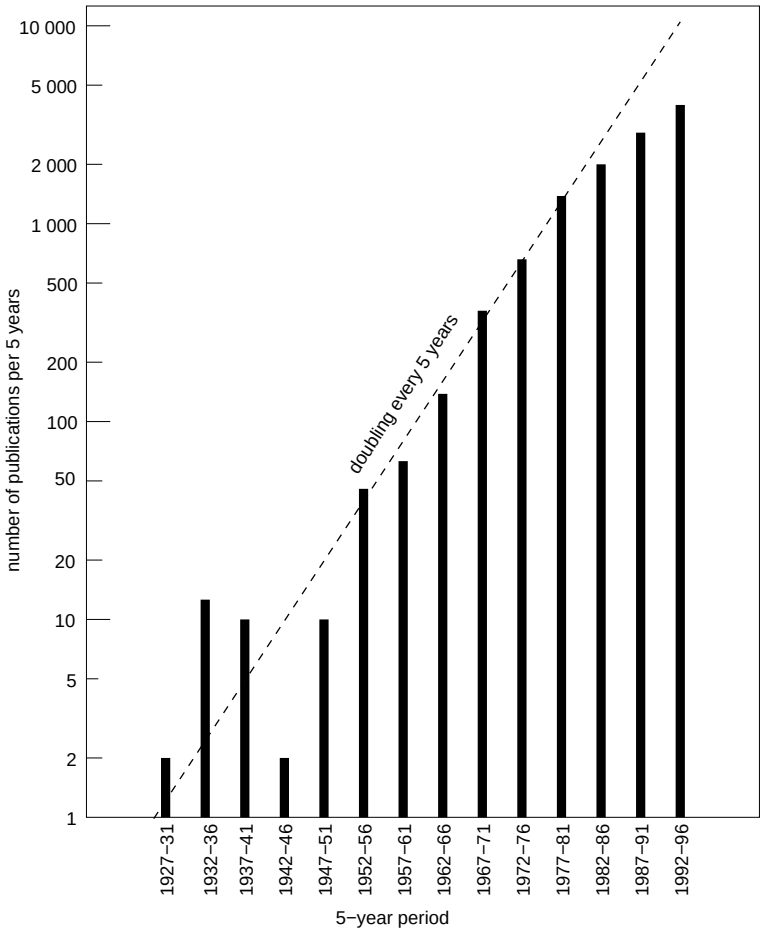


Figure 8. Five-year cumulants of ANVC publications, based on the author's data files.

catastrophic consequences.

The general interest in active control of noise and vibration has been steadily increasing, a fact which can also be concluded from the growing number of textbooks, special conferences and journal papers per year. Based on the author's collection of more than 12 000 references on active control of sound and vibration [219,220], Fig. 8 shows a histogram of the number of publications grouped in five-year periods. An exponential increase is observed from the 1950s through the early 1980s with doubling every 5 years, followed by further growth at reduced pace (approximately doubling in ten years). There are nearly twice as many papers on active vibration control than on active sound field control, and there are about 7% patent applications. This is relatively high for a research topic and proves the considerable commercial interest.

References

[1] Lord Rayleigh, *The Theory of Sound*, Vol. II, Chapter XIV, § 282: 'Two Sources of Like Pitch; Points of Silence; Experimental Methods' (MacMillan & Co, London, 1st ed. 1878): pp. 104–106; 2nd ed. (1894/96) and Reprints (Dover, New York): pp. 116–118).

[2] H. Coanda, 'Procédé de protection contre les bruits', French Patent FR 722.274, filed: Oct. 21, 1930, patented: Dec. 29, 1931.

[3] H. Coanda, 'Procédé et dispositif de protection contre les bruits', French Patent FR 762.121, filed: Dec. 31, 1932, patented: Jan. 18, 1934.

[4] P. Lueg, 'Verfahren zur Dämpfung von Schallschwingungen', German Patent No. 655 508, filed: Jan. 27,1933, patented: Dec. 30, 1937.

[5] P. Lueg, 'Process of Silencing Sound Oscillations', U.S. Patent US 2,043,416, filed: March 8, 1934, patented: June 9, 1936.

[6] H. F. Olson, 'Electronic Sound Absorber', U.S. Patent US 2,983,790, filed: April 30, 1953, patented: May 9, 1961.

[7] H. F. Olson, 'Electronic Control of Noise, Vibration, and Reverberation', J. Acoust. Soc. Am. **28**, 966–972 (1956).

[8] L. Schreiber, 'Schallschutz durch "Antilärmquellen"? Wirkliche Lärmminderung nur durch Vernichtung der Schallenergie', VDI-Nachrichten Nr. 15, p. 13 (April 14, 1971).

[9] M. J. M. Jessel, 'La question des absorbeurs actifs', Revue d'Acoustique **5**, No. 18, pp. 37–42 (1972).

[10] G. Mangiante, 'Active Sound Absorption', J. Acoust. Soc. Am. **61**, 1516–1523 (1977).

[11] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic Press, London, 1992).

[12] G. Mangiante and J. P. Vian, 'Application du principe de Huygens aux absorbeurs acoustiques actifs. II: Approximations du principe de Huygens', Acustica **37**, 175–182 (1977).

[13] J. Piraux and S. Mazzanti, 'Broadband active noise attenuation in three-dimensional space', in *Proc. Internoise 85, München, Germany* (1985), pp. 485–488.

[14] H.-J. Lee et al., 'An Active Noise Control System for Controlling Humming Noise Generated by a Transformer', in *Proc. Internoise 97, Budapest, Aug. 25–27, 1997*, Vol. I, pp. 517–520.

[15] S. Uosukainen, 'Active sound scatterers based on the JMC method', J. Sound Vib. **267**, 979–1005 (2003).

[16] S. Takane and T. Sone, 'A new theory for active suppression of reflected sound waves from the walls based on Kirchhoff-Helmholtz boundary integral equation', Acoustical Science and Technology **25**, 37–44 (2004).

[17] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall Inc., Englewood Cliffs, NJ, US, 1985).

[18] D. Guicking and K. Karcher, 'Active Impedance Control for One-Dimensional Sound', ASME J. Vib. Acoust. Stress Rel. in Design **106**, 393–396 (1984).

[19] M. A. Swinbanks, 'The active control of sound propagation in long ducts', J. Sound Vib. **27**, 411–436 (1973).

[20] D. Guicking and H. Freienstein, 'Broadband Active Sound Absorption in Ducts with Thinned Loudspeaker Arrays', in *Proc. ACTIVE 95, Newport Beach, CA, US, July 6–8, 1995*, pp. 371–382.

[21] L. J. Eriksson, 'Active Attenuation System with On-line Modeling of Speaker, Error Path and Feedback Path', U.S. Patent US 4,677,676, filed: Dec. 11, 1986, patented: June 30, 1987.

[22] P. L. Feintuch, 'An Adaptive Recursive LMS Filter', Proc. IEEE **64**, 1622–1624 (1976).

Comments: **65**, 1399 and 1402 (1977).
[23] W. von Heesen, 'Practical Experience with an Active Noise Control Installation in the Exhaust Gas Line of a Co-Generator Plant Engine', ACUSTICA – acta acustica **82**, Suppl. 1, p. S 195 (1996).
[24] S. Deus, 'Aktive Schalldämpfung im Ansaugkanal von Gebläsen', in *Fortschritte der Akustik – DAGA 98 (Zürich 1998)*, (DEGA e.V., Oldenburg, Germany, 1998), pp. 686–687.
[25] J. Hansen, 'Eine anwendungsreife Lösung für die aktive Minderung von Abgasgeräuschen industrieller Dieselmotoren und Drehkolbenpumpen', VDI Bericht Nr. 1491, 199 (1999).
[26] C. H. Hansen et al., 'Practical implementation of an active noise control system in a hot exhaust stack', Acoustics 1996 – Australian Acoustical Society Annual Conference, Brisbane, AU, Nov. 13–15, 1996.
[27] L. J. Eriksson et al., 'Active Acoustic Attenuation System for Higher Order Mode Non-Uniform Sound Field in a Duct', U.S. Patent US 4,815,139, filed: March 16, 1988, patented: March 21, 1989.
[28] G. V. Moustakides, 'Correcting the Instability Due to Finite Precision of the Fast Kalman Identification Algorithms', Signal Processing **18**, 33–42 (1989).
[29] R. Schirmacher and D. Guicking, 'Theory and implementation of a broadband active noise control system using a fast RLS algorithm', Acta Acustica **2**, 291–300 (1994).
[30] S. R. Popovich, 'Fast Adapting Control System and Method', U.S. Patent US 5,602,929, filed: Jan. 30, 1995, patented: Feb. 11, 1997.
[31] M. O. Tokhi and R. R. Leitch, *Active Noise Control*, The Oxford Engineering Sciences Series, Vol. 29, (Clarendon Press, Oxford, UK, 1992).
[32] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems. Algorithms and DSP Implementations*, Wiley Series in Telecommunication and Signal Processing (John Wiley & Sons, New York, 1996).
[33] R. Schirmacher, 'Schnelle Algorithmen für adaptive IIR-Filter und ihre Anwendung in der aktiven Schallfeldbeeinflussung', Doctoral Thesis, Göttingen, Germany, 1995. Abstract: ACUSTICA – acta acustica **82**, 384 (1996).
[34] W. S. Levine, *The Control Handbook* (CRC Press/IEEE Press, 1995).
[35] S. O. R. Moheimani et al., *Spatial Control of Vibration: Theory and Experiments* (World Scientific Press (UK), London, 2003).
[36] B. Widrow, 'Adaptive Noise Cancelling: Principles and Applications', Proc. IEEE **63**, 1692–1716 (1975).
[37] J. M. McCool et al., 'Adaptive detector', U.S. Patent US 4,243,935, filed: May 18, 1979, patented: Jan. 6, 1981.
[38] M. H. Silverberg et al., 'Outbound Noise Cancellation for Telephonic Handset', U.S. Patent US 5,406,622, filed: Sept. 2, 1993, patented: April 11, 1995'.
[39] B. Widrow, 'Seismic Exploration Method and Apparatus for Cancelling Interference from Seismic Vibration Source', U.S. Patent US 4,556,962, filed: April 21, 1983, patented: Dec. 3, 1985.
[40] B. Widrow, 'ECG Enhancement by Adaptive Cancellation of Electrosurgical Interference', U.S. Patent US 4,537,200, filed: July 7, 1983, patented: Aug. 27, 1985.
[41] T. R. Harley, 'Active Noise Control Stethoscope', U.S. Patent US 5,610,987, first filed: Aug. 16, 1993, patented: March 11, 1997.
[42] H. Ding, 'An adaptive speech enhancement method for siren noise cancellation', Applied Acoustics **65**, 385–399 (2004).
[43] J. Vanden Berghe and J. Wouters, 'An adaptive noise canceller for hearing aids using two nearby microphones', J. Acoust. Soc. Am. **103**, 3621–3626 (1998).

[44] Falcke, H., 'Lofar und die Epoche der Reionisation', Spektr. d. Wiss., Jan. 2007, 51.
[45] G. P. Eatwell et al., 'Piezo Speaker for Improved Passenger Cabin Audio Systems', International Patent Application WO 97/17818 A1, published: May 15, 1997, priority (US): Sept. 25, 1995.
[46] H.-J. Raida and O. Bschorr, 'Gerichteter Stabstrahler', German Patent DE 196 48 986 C1, filed: Nov. 26, 1996, patented: April 9, 1998.
[47] M. A. Daniels, 'Loudspeaker Phase Distortion Control Using Velocity Feedback', U.S. Patent US 5,771,300, filed: Sept. 25, 1996, patented: June 23, 1998.
[48] C. Carme et al., 'Haut-parleur linéaire', French Patent Application FR 2 766 650 A1, filed: July 23, 1997, published: Jan. 29, 1999.
[49] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill Book Co., New York, 1968).
[50] M. A. Swinbanks, 'The active control of noise and vibration and some applications in industry', Proc. IMechE **198A**, No. 13, pp. 281–288 (1984).
[51] G. B. B. Chaplin, 'Anti-sound — The Essex breakthrough', Chartered Mechanical Engineer (CME) **30**, 41–47 (Jan. 1983).
[52] F. Lehringer and G. Zintel, 'Aktive Pegelminderung bei Abgasanlagen von Kraftfahrzeugen', Haus der Technik, Essen, Germany, Information material to the meeting on "Aktive Lärmbekämpfung und Schwingungsabwehr", Feb. 21, 1995 (12 pp.).
[53] J. Krüger et al., 'Aktive Abgas-Schalldämpfer für PKW – Chancen und Risiken', in *Fortschritte der Akustik – DAGA 2005, München, DE, March 14–17* (2005), pp. 21–22.
[54] Kh. Eghtesadi et al., 'Industrial Applications of Active Noise Control', in *Proc. NOISE-93, International Noise and Vibration Control Conference, St. Petersburg, Russia, May 31 – June 3, 1993*, Vol. 2, pp. 147–152.
[55] W. Niehoff, 'A simple optical microphone', J. Acoust. Soc. Am. **105**, 1052 (1999) (Abstract). Papers presented at the Joint ASA/EAA/DAGA Meeting, Berlin, Germany, March 14–19, 1999. Collected Papers on CD ROM: paper 2pEA2 (4 pp.) (Title: 'The Optical Microphone Using Glass Fibre Technology').
[56] G. K. Behler and T. Lentz, 'A Tube Headphone without Metallic Parts to be Used for Functional NMR', in *Proc. CFA/DAGA '04, Congrès Joint 7ème Congrès Français d'Acoustique / 30. Deutsche Jahrestagung für Akustik, Strasbourg, FR, March 22–25, 2004*, pp. 1035–1036.
[57] J. Tyrrell, 'Optical microphone aids MRI scans', Opto & Laser Europe (OLE), No. 126, p. 11 (March 2005).
[58] J. Qiu J. and Tani, 'Vibration control of a cylindrical shell used in MRI equipment', Smart Materials and Structures **4**, A75–A81 (1995).
[59] T. Nestorović Trajkov et al., 'Vibration control of a funnel-shaped shell structure with distributed piezoelectric actuators and sensors', Smart Materials and Structures **15**, 1119–1132 (2006).
[60] A. V. Bykhovsky, 'Sposob polawlenija schuma w sluchowom organe (Method for noise reduction in the ear)', UdSSR Patent SU 133 631, filed: Aug. 24, 1949, published: Invention Bulletin No. 22, 1960.
[61] R. L. McKinley, 'Development of Active Noise Reduction Earcups for Military Applications', ASME Winter Annual Meeting, Anaheim, CA, US, Dec. 1986, Session NCA-8B.
[62] I. Veit, 'Gehörschutz-Kopfhörer. Elektronik kontra Lärm', Funkschau **60**, No. 23, 50–52 (1988).
[63] L. R. Ray et al., 'Hybrid feedforward–feedback active noise reduction for hearing protection and communication', J. Acoust. Soc. Am. **120**, 2026–2036 (2006).

[64] M. R. Schroeder, D. Gottlob and K. F. Siebrasse, 'Comparative Study of European Concert Halls. Correlation of Subjective Preference with Geometric and Acoustic Parameters', J. Acoust. Soc. Am. **56**, 1195–1201 (1974).
[65] S. M. Kuo et al., 'Acoustic Noise and Echo Cancellation Microphone System for Videoconferencing', IEEE Transactions on Consumer Electronics **41**, 1150–1158 (1995).
[66] E. Hänsler, 'The hands-free telephone problem – An annotated bibliography', Signal Processing **27**, 259–271 (1992).
[67] A. Ortega et al, 'Speech Reinforcement System for Car Cabin Communications', IEEE Trans. Speech and Audio Processing **13**, 917–929 (2005).
[68] T. Gänsler and J. Benesty, 'New Insights to the Stereophonic Acoustic Echo Cancellation Problem and an Adaptive Nonlinearity Solution', IEEE Trans. Speech and Audio Processing **10**, 257–267 (2002).
[69] A. W. H. Khong and P. A. Naylor, 'Stereophonic Acoustic Echo Cancellation Employing Selective-Tap Adaptive Algorithms', IEEE Trans. Audio, Speech, and Language Processing **14**, 785–796 (2006).
[70] S. Gustafsson et al., 'A Psychoacoustic Approach to Combined Acoustic Echo Cancellation and Noise Reduction', IEEE Trans. Speech and Audio Processing **10**, 245–256 (2002).
[71] O. Hoshuyama and R. A. Goubran, 'A new adaptation algorithm for echo cancellation in fast changing enviroments', in *Proc. 18th International Congress on Acoustics (ICA 2004), Kyoto, JP, April 4–9, 2004*, Paper Th.P1.13, pp. IV-3147 – IV-3150.
[72] A. Cichocki and Shun-ichi Amari, *Adaptive Blind Signal and Image Processing* (John Wiley & Sons, New York, 2006).
[73] M. M. Sondhi, 'Closed Loop Vibration Echo Canceller Using Generalized Filter Networks', U.S. Patent US 3,499,999, filed: Oct. 31, 1966, patented: March 10, 1970.
[74] E. Herter and W. Lörcher, *Nachrichtentechnik* (Carl Hanser-Verlag, München and Wien, 5th Edition, 1990), Section 7.5.6.
[75] G. P. Eatwell, 'The Use of the Silentseat in Aircraft Cabins', in *Proceedings of the 1st Conference on Recent Advances in Active Control of Sound and Vibration* edited by C. A. Rogers and C. R. Fuller (Blacksburg, VA, US, April 15–17, 1991) pp. 302–310.
[76] C. G. Hutchens and S. A. Morris, 'Method for Acoustic Reverberation Removal', U.S. Patent US 4,796,237, filed: Jan. 28, 1987, patented: Jan. 3, 1989.
[77] D. H. Gilbert, 'Echo Cancellation System', U.S. Patent US 4,875,372, filed: May 3, 1988, patented: Oct. 24, 1989.
[78] D. Hassler, 'Apparatus and Method for Suppressing Reflections at an Ultrasound Transducer', U.S. Patent US 5,245,586, patented: Sept. 14, 1993, priority (EP): Nov. 15, 1991.
[79] M. Wenzel, 'Untersuchungen zur breitbandigen Messung und Regelung der akustischen Wandimpedanz an einer aktiven Schallwand mit adaptiven Filtern', Doctoral Thesis, Göttingen, Germany, 1992.
[80] D. Guicking, K. Karcher and M. Rollwage, 'Active control of the acoustic reflection coefficient at low frequencies', in *Internoise 83, International Congress on Noise Control Engineering, Edinburgh, UK, July 13–15, 1983*, pp. 419–422.
[81] D. Guicking, J. Melcher and R. Wimmel, 'Active Impedance Control in Mechanical Systems', Acustica **69**, 39–52 (1989).
[82] S. Ise, 'Theory of Acoustic Impedance Control for Active Noise Control', in *Proc. Internoise 94, International Congress on Noise Control Engineering, Yokohama, JP, Aug. 29–31, 1994*, pp. 1339–1342.
[83] O. Lacour et al., 'Preliminary experiments on noise reduction in cavities using active impedance changes', J. Sound Vib. **230**, 69–99 (2000).

[84] J. Melcher, 'Adaptive Impedanzregelung an strukturmechanischen Systemen', Doctoral Thesis, University of Magdeburg, DE, March 2001 (Shaker Verlag, Aachen, 2001).
[85] M.-A. Galland, 'Active control and porous materials', Forum Acusticum, Sevilla, ES, Sept. 16–20, 2002, Proceedings on CD ROM: file ele01003.pdf (6 pp.).
[86] M.-A. Galland, 'Hybrid passive/active absorbers for flow ducts', Applied Acoustics **66**, 691–708 (2005).
[87] S. J. Elliott and P. A. Nelson, 'Multichannel active sound control using adaptive filtering', in *Proc. ICASSP '88* (1988), Paper A3.4, pp. 2590–2593.
[88] W. Böhm, 'Untersuchungen zur breitbandig wirksamen aktiven Kompensation instationär angeregter Schallfelder', Doctoral Thesis, Göttingen, Germany, 1992.
[89] M. Bronzel, 'Aktive Schallfeldbeeinflussung nicht-stationärer Schallfelder mit adaptiven Digitalfiltern', Doctoral Thesis, Göttingen, Germany, 1993'.
[90] R. Freymann, 'Von der Pegelakustik zum Sounddesign', in *Fortschritte der Akustik – DAGA '96, DEGA e. V.* (Oldenburg, Germany, 1996), pp. 32–42.
[91] J. Scheuren, U. Widmann and J. Winkler, 'Active Noise Control and Sound Quality Design in Motor Vehicles', SAE Technical Paper Series, No. 1999-01-1846. Proc. of the 1999 Noise and Vibration Conference and Exposition, Traverse City, MI, US, May 17–20, 1999.
[92] A. González, 'Sound quality of low-frequency and car engine noises after active noise control', J. Sound Vib. **265**, 663–679 (2003).
[93] L. E. Rees and S. J. Elliott, 'Adaptive Algorithms for Active Sound-Profiling', IEEE Trans. Audio, Speech, and Language Processing **14**, 711–719 (2006).
[94] C. R. Fuller, 'Analytical model for investigation of interior noise characteristics in aircraft with multiple propellers including synchrophasing', J. Sound Vib. **109**, 141–156 (1986).
[95] S. Johansson et al., 'Performance of a Multiple Versus a Single Reference MIMO ANC Algorithm Based on a Dornier 328 Test Data Set', in *Proceedings of ACTIVE 97, the 1997 International Symposium on Active Control of Sound and Vibration, Budapest, HU, Aug. 21–23, 1997*, pp. 521–528.
[96] M. Kallergis, 'Experimental Results on Propeller Noise Attenuation Using an 'Active Noise Control' Technique', in *Proc. 14th DGLR/AIAA Aeroacoustics Conference, Aachen, Germany, May 11–14, 1992*, pp. 907–918, AIAA Paper 92-02-155.
[97] I. Veit, 'Anordnung und Verfahren zur aktiven Reduzierung von Reifenschwingungen', German Patent DE 197 23 516 C1, filed: June 5, 1997, patented: Oct. 29, 1998.
[98] W. B. Conover and W. F. M. Gray, 'Noise Reducing System for Transformers', U.S. Patent US 2,776,020, filed: Feb. 9, 1955, patented: Jan. 1, 1957.
[99] G. B. B. Chaplin, 'The Cancelling of Vibrations Transmitted through a Fluid in a Containing Vessel', International Patent Application WO 81/01479 A1, published: May 28, 1981, priority (GB): Nov. 10, 1979.
[100] Y. Hori et al., 'Vibration/Noise Reduction Device for Electrical Apparatus', U.S. Patent US 4,435,751, patented: March 6, 1984, priority (JP): July 3, 1980.
[101] W. Gossman and G. P. Eatwell, 'Active High Transmission Loss Panel', U.S. Patent US 5,315,661, filed: Aug. 12, 1992, patented: May 24, 1994.
[102] O. L. Angevine, 'Active Systems for Attenuation of Noise', Int. J. of Active Control **1**, 65–78 (1995).
[103] Y. Ai, X. Qiu and C. H. Hansen, 'Minimizing wind effects on active control systems for attenuating outdoor transformer noise', Noise Control Engineering J. **48**, 130–135 (2000).
[104] X. Qiu et al., 'A waveform synthesis algorithm for active control of transformer noise:

Implementation', Applied Acoustics **63**, 467–479 (2002).
[105] S. Ise, 'Basic study on active noise barrier', J. Acoust. Soc. Japan (E) **12**, 299–306 (1991).
[106] T. Nakashima and S. Ise, 'Active noise barrier for far field noise reduction', in *Proc. 18th International Congress on Acoustics (ICA 2004), Kyoto, JP, April 4–9*, Paper We4.D.4, pp. III-2161 – III-2164.
[107] A. P. Berkhoff, 'Control strategies for active noise barriers using near-field error sensing', J. Acoust. Soc. Am. **118**, 1469–1479 (2005).
[108] A. Mallock, 'A Method of Preventing Vibration in Certain Classes of Steamships', Trans. Inst. Naval Architects **47**, 227–230 (1905).
[109] H. Hort, 'Beschreibung und Versuchsergebnisse ausgeführter Schiffsstabilisierungsanlagen', Jahrb. Schiffbautechn. Ges. **35**, 292–312 (1934).
[110] J. F. Allan, 'The Stabilization of Ships by Activated Fins', Trans. Inst. Naval Architects **87**, 123–159 (1945).
[111] A. Vang, 'Vibration Dampening', U.S. Patent US 2,361,071, filed: Sept. 23, 1942, patented: Oct. 24, 1944.
[112] W. P. Mason, 'Piezoelectric Damping Means for Mechanical Vibrations', U.S. Patent US 2,443,417, filed: March 29, 1945, patented: June 15, 1948.
[113] Anon., 'Perfectionnements apportés aux paliers pour corps tournants, notamment pour ensembles devant tourner à l'interieur d'une enceinte étanche', French Patent FR 1 186 527, filed: Nov. 18, 1957, patented: Feb. 23, 1959.
[114] J. D. Leatherwood et al., 'Active Vibration Isolator for Flexible Bodies', U.S. Patent US 3,566,993, filed: March 26, 1969, patented: March 2, 1971.
[115] M. Mano, 'Ship Design Considerations for Minimal Vibration', in *Proc. Ship Technology and Research (STAR) 10th Symposium of the Society of Naval Architects and Marine Engineers (SNAME)* (1985), pp. 143–156.
[116] L. Meirovitch, *Dynamics and Control of Structures* (John Wiley & Sons, New York, 1990).
[117] L. Gaul, 'Aktive Beeinflussung von Fügestellen in mechanischen Konstruktionselementen und Strukturen', German Patent Application DE 197 02 518 A1, filed: Jan. 24, 1997, published: June 12, 1997.
[118] P. Buaka et al., 'Optimal energy dissipation in a semi-active friction device', J. Acoust. Soc. Am. **117**, 2602 (2005) (Abstract).
[119] D. Tang et al., 'Study of airfoil gust response alleviation using an electro-magnetic dry friction damper. Part 1: Theory' and '... Part 2: Experiment', J. Sound Vib. **269**, 853–874 and 875–897 (2004).
[120] I. Tuzcu and L. Meirovitch, 'Control of flying flexible aircraft using control surfaces and dispersed piezoelectric actuators', Smart Materials and Structures **15**, 893–903 (2006).
[121] R. Freymann, 'Dynamic Interactions Between Active Control Systems and a Flexible Aircraft Structure', in *Proc. 27th AIAA/ASME/SAE SDM Conference* (1986), pp. 517–524, AIAA Paper 86-0960.
[122] H. Heller et al., 'Helicopter Rotor Noise Research in Aeroacoustic Wind Tunnels – State of the Art and Perspectives', in *Proc. NOISE-93, International Noise and Vibration Control Conference, St. Petersburg, May 31 – June 3, 1993*, Vol. 4, pp. 39–60.
[123] Y. Chen et al., 'Experimental evaluation of the Smart Spring for helicopter vibration suppression through blade root impedance control', Smart Materials and Structures **14**, 1066–1074 (2005).
[124] R. L. Clark and G. P. Gibbs, 'A novel approach to feedforward higher-harmonic control', J. Acoust. Soc. Am. **96**, 926–936 (1994).

[125] D. Heyland et al., 'The adaptive wing project (DLR): Survey on targets and recent results from active/adaptive structures viewpoint', in *Proc. Tenth International Conference on Adaptive Structures and Technologies (ICAST '99), Paris, FR, Oct. 11–13, 1999* (2000), pp. 178–185.

[126] J. Melcher and A. Büter, 'Adaptive Structures Technology for Structural Acoustic Problem', in *1st Joint CEAS/AIAA Aeronautics Conference (16th AIAA Aeroacoustics Conference), München, Germany, June 12–15, 1995*, pp. 1213–1220.

[127] C. R. Fuller, S. J. Elliott and P. A. Nelson, *Active Control of Vibration* (Academic Press, London, 1995).

[128] H.-P. Monner, E. J. Breitbach and H. Hanselka, 'Recent Results and Future of the German Major Project Adaptronics', in *Proc. Thirteenth International Conference on Adaptive Structures and Technologies (ICAST 2002), Potsdam, DE, Oct. 7–9, 2002*, edited by E. J. Breitbach et al., pp. 259–269.

[129] T. L. Lagö, 'Industry Overview of Active Control Methods and Applications', Adaptronic Congress, Berlin/Potsdam 2002, Paper 19 (also on CD ROM).

[130] D. L. Edberg and A. H. von Flotow, 'Progress Toward a Flight Demonstration of Microgravity Isolation of Transient Events', World Space Congress, 43rd Congress of the International Astronautical Federation, Paper IAF-92-0781.

[131] G. J. Stein, 'A Driver's Seat with Active Suspension of Electropneumatic Type', ASME J. of Vibration and Acoustics **119**, 230–235 (1997).

[132] G. Kim and R. Singh, 'A Study of Passive and Adaptive Hydraulic Engine Mount Systems with Emphasis on Non-Linear Characteristics', J. Sound Vib. **179**, 427–453 (1995).

[133] L. Maack and J. Stäbler, 'Active Body Control in Production', dSPACE NEWS, Fall 2000, pp. 2–3.

[134] S. Asiri, A. M. Baz and D. Pines, 'Active periodic struts for a gearbox support system', Smart Materials and Structures **15**, 1707–1714 (2006).

[135] J. Wang et al., 'Active vibration control of a plate-like structure with discontinuous boundary conditions', Smart Materials and Structures **15**, N51–N60 (2006).

[136] S. Morishita and J. Mitsui, 'An Electronically Controlled Engine Mount Using Electro-Rheological Fluid', SAE Special Publication **936**, 97–103 (1992).

[137] Anon., 'Intelligent Solutions. Engineering materials ease the burden on the environment', in: *Chemistry with Chlorine*, Bayer AG, Public Relations Department (Leverkusen, Germany, April 1995).

[138] M. Taki et al., 'Sound Attenuating System', U.S. Patent US 5,347,585, filed: Sept. 10, 1991, patented: Sept. 13, 1994.

[139] S. J. Dyke et al., 'Modeling and control of magnetorheological dampers for seismic response reduction', Smart Materials and Structures **5**, 565–575 (1996).

[140] J. Wang and G. Meng, 'Experimental study on stability of an MR fluid damper-rotor-journal bearing system', J. Sound Vib. **262**, 999–1007 (2003).

[141] Hyun-Ung Oh, 'Experimental demonstration of an improved magneto-rheological fluid damper for suppression of vibration of a space flexible structure', Smart Materials and Structures **13**, 1238–1244 (2004).

[142] S. Sassi et al., 'An innovative magnetorheological damper for automotive suspension: from design to experimental characterization', Smart Materials and Structures **14**, 811–822 (2005).

[143] L. Chen and C. H. Hansen, 'Active Vibration Control of a Magnetorheological Sandwich Beam', in *Proc. Acoustics 2005 – Acoustics in a Changing Environment, Australian Acoustical Society Annual Conference, Busselton, AU, Nov. 9–11, 2005*, pp. 93–98.

[144] Q. Zhou et al., 'Semi-active control of three-dimensional vibrations of an inclined sag cable with magnetorheological dampers', J. Sound Vib. **296**, 1–22 (2006).
[145] H.-X. Deng et al., 'Development of an adaptive tuned vibration absorber with magnetorheological elastomer', Smart Materials and Structures **15**, N111–N116 (2006).
[146] C. Carmignani et al., 'Design of a novel magneto-rheological squeeze-film damper', Smart Materials and Structures **15**, 164–170 (2006).
[147] N. R. Petersen, 'Design of Large Scale Tuned Mass Dampers', in *Structural Control*, edited by H. H. E. Leipholz (North-Holland Publ. Co., 1980), pp. 581–596.
[148] J. C. H. Chang and T. T. Soong, 'The use of aerodynamic appendages for tall building control', in *Structural Control*, edited by H. H. E. Leipholz (North-Holland Publ. Co., 1980), pp. 199–210.
[149] M. Izumi, 'Control of Structural Vibration – Past, Present and Future', in *Proc. International Symposium on Active Control of Sound and Vibration, Tokyo, April 1991*, pp. 195–200.
[150] Q. S. Li et al., 'Combinatorial optimal design of number and positions of actuators in actively controlled structures using genetic algorithms', J. Sound Vib. **270**, 611–624 (2004).
[151] T. T. Soong and H. G. Natke, 'From Active Control to Active Structures', VDI-Berichte **695**, 1–18 (1988).
[152] K. Hanahara and Y. Tada, 'Dynamics of Geometry Adaptive Truss with Wire Member Actuators', in *Proc. 13th International Conference on Adaptive Structures and Technologies (ICAST 2002), Potsdam, DE, Oct. 7–9, 2002*, edited by E. J. Breitbach et al. (2004), pp. 403–412.
[153] F. Merkle, 'Aktive und adaptive Optik in der Astronomie. Neue Technologien für zukünftige Großteleskope', Phys. Blätter **44**, 439–446 (1988).
[154] B. Schwarzschild, 'First of the Twin 10-Meter Keck Telescopes Starts Doing Astronomy', Phys. Today **46** (10), 17–18 (1993).
[155] R. Q. Fugate, 'Measurement of atmospheric wavefront distortion using scattered light from a laser guide-star', Nature **353**, 144–146 (1991).
[156] Ch. Baranec et al., 'Concept for a laser guide beacon Shack-Hartmann wave-front sensor with dynamically steered subapertures', Opt. Lett. **30**, 693–695 (2005).
[157] G. P. Collins , 'Making Stars to See Stars: DOD Adaptive Optics Work is Declassified', Phys. Today **45** (2), 17–21 (1992).
[158] B. Schwarzschild, 'Adaptive Optics at the New 8-Meter Gemini Telescope', Phys. Today **52** (9), 23 (1999).
[159] R. Gilmozzi, 'Riesenteleskope der Zukunft', Spektrum der Wissenschaft, 28–36 (Aug. 2006).
[160] Q. Yang et al., 'Wavefront correction with high-order curvature adaptive optics systems', J. Opt. Soc. Am. A **23**, 1375–1381 (2006).
[161] J. Bell, "Adaptive optics clears the view for industry', Opto & Laser Europe (OLE) No. 45, pp. 17–20, (Nov. 1997).
[162] H. Baumhacker, 'Correction of strong phase and amplitude modulations by two deformable mirrors in a multistaged Ti:sapphire laser', Opt. Lett. **27**, 1570–1572 (2002).
[163] D. Burns, 'University of Strathclyde: intracavity adaptive-optic control of lasers', Opto & Laser Europe, No. 126, p. 27, (March 2005).
[164] M. J. Booth et al., 'New modal wave-front sensor: application to adaptive confocal fluorescence microscopy and two-photon excitation fluorescence microscopy', J. Opt. Soc. Am. A **19**, 2112–2120 (2002).
[165] H. Hemmati and Y. Chen, 'Active optical compensation of low-quality optical system aberrations', Opt. Lett. **31**, 1630–1632 (2006).

[166] S. Zommer et al., 'Simulated annealing in ocular adaptive optics', Opt. Lett. **31**, 939–941 (2006).
[167] C. R. Vogel and Q. Yang, 'Modeling, simulation, and open-loop control of a continuous facesheet MEMS deformable mirror', J. Optical Soc. Am. A **23**, 1074–1081 (2006).
[168] A. Greenaway and J. Burnett, *Industrial and Medical Applications of Adaptive Optics* (Institute of Physics (IOP) Publishing Ltd., Bristol, GB, 2004).
[169] R. K. Tyson, *Principles of Adaptive Optics.* (Academic Press, San Diego, CA, US, 1991).
[170] V. P. Lukin, *Atmospheric Adaptive Optics* (SPIE Press Vol. PM23, Feb. 1996).
[171] J. W. Hardy, *Adaptive Optics for Astronomical Telescopes* (Oxford University Press, New York, 1998).
[172] F. Roddier (ed.), *Adaptive Optics in Astronomy* (Cambridge University Press, New York, 1999).
[173] D. T. Busher et al., *Adaptive Optics* (John Wiley & Sons, Ltd., New York, 2002).
[174] C. R. Fuller et al., 'Active Control of Sound Radiation from a Vibrating Rectangular Panel by Sound Sources and Vibration Inputs: An Experimental Comparison', J. Sound Vib. **145**, 195–215 (1991).
[175] R. L. Clark and C. R. Fuller, 'Active Structural Acoustic Control with Adaptive Structures Including Wavenumber Considerations', in *Proc. of the 1st Conference on Recent Advances in Active Control of Sound and Vibration*, edited by C. A. Rogers and C. R. Fuller, (Blacksburg, VA, US, April 1991) pp. 507–524.
[176] C. R. Fuller, S. J. Elliott and P. A. Nelson, *Active Control of Vibration* (Academic Press, London, 1995).
[177] M. R. F. Kidner and R. I. Wright, 'Global control of sound radiation from a plate using several adaptive vibration neutralisers with local control schemes', in *Proc. Acoustics 2005 – Acoustics in a Changing Environment, Australian Acoustical Society Annual Conference, Busselton, AU, Nov. 9–11*, pp. 55–60.
[178] A. Jakob and M. Möser, 'Parameter Study with a Modal Model for Actively Controlled Double-Glazed Windows', Acta Acustica united with Acustica **90**, 467–480 (2004).
[179] J. P. Carneal and C. R. Fuller, 'An analytical and experimental investigation of active structural acoustic control of noise transmission through double panel systems', J. Sound Vib. **272**, 749–771 (2004).
[180] S. Akishita et al., 'Active Modular Panel System for Insulating Floor Impulse Noise', in *Proc. 18th International Congress on Acoustics (ICA 2004), Kyoto, JP, April 4–9, 2004*, Paper We4.D.5, pp. III-2165 – III-2168.
[181] S.-B. Choi, 'Active structural acoustic control of a smart plate featuring piezoelectric actuators', J. Sound Vib. **294**, 421–429 (2006).
[182] J. Kim and J.-H. Kim, 'Multimode shunt damping of piezoelectric smart panel for noise reduction', J. Acoust. Soc. Am. **116**, 942–948 (2004).
[183] D. Guyomar et al., 'Wave reflection and transmission reduction using a piezoelectric semipassive nonlinear technique', J. Acoust. Soc. Am. **119**, 285–298 (2006).
[184] G. A. Lesieutre et al., 'Damping as a result of piezoelectric energy harvesting', J. Sound Vib. **269**, 991-1001 (2004).
[185] S. Jiang, 'Performance of a piezoelectric bimorph for scavenging vibration energy', Smart Materials and Structures **14**, 769–774 (2005).
[186] J. C. Ji and C. H. Hansen, 'Hopf Bifurcation of a Magnetic Bearing System with Time Delay', ASME J. Vib. Acoust. **127**, 362–369 (2005).
[187] T. Kapitaniak, *Chaos for Engineers* (Springer-Verlag, Berlin, 1998).
[188] H. G. Schuster (ed.), *Handbook of Chaos Control* (Wiley-VCH, Weinheim, 1999).
[189] A. Hübler and E. Lüscher, 'Resonant stimulation and control of nonlinear oscillators',

Die Naturwissenschaften **76**, 67–69 (1989).
[190] E. Ott, C. Grebogi and Y. A. Yorke, 'Controlling chaos', Phys. Rev. Lett. **64**, 1196–1199 (1990).
[191] K. Pyragas, 'Continuous control of chaos by self-controlling feedback', Phys. Lett. A **70**, 421–428 (1992).
[192] H.-T. Yau et al., 'Sliding mode control of chaotic systems with uncertainties', Int. J. Bifurcation and Chaos **10**, 1139–1147 (2000).
[193] F. Schürer, 'Zur Theorie des Balancierens', Math. Nachr. **1**, 295–331 (1948).
[194] W. L. Ditto et al., 'Control of human atrial fibrillation', Int. J. Bifurcation and Chaos **10**, 593–601 (2000).
[195] L. Kocarev and U. Parlitz, 'General approach for chaotic synchronization with application to communication', Phys. Rev. Lett. **74**, 5028–5031 (1995).
[196] A. Ahlborn and U. Parlitz, 'Multiple Delay Feedback Control', to appear in the forthcoming 2nd edition of Ref. [188].
[197] G. Chen et al., 'Bifurcation control: Theories, methods, and applications', Int. J. Bifurcation and Chaos **10**, 511–548 (2000).
[198] X. F. Wang and G. Chen, 'Chaotification via arbitrarily small feedback controls: Theory, method, and applications', Int. J. Bifurcation and Chaos **10**, 549–570 (2000).
[199] U. Parlitz, 'Complex Dynamics of Nonlinear Systems', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[200] P. J. Dines, 'Active control of flame noise', Ph.D. Thesis, Cambridge, England (1984).
[201] H. W. Liepmann and D. M. Nosenchuk, 'Active Control of Laminar–Turbulent Transition', J. Fluid Dynamics **118**, 201–204 (1982).
[202] D. Egelhof, 'Einrichtung zur Schwingungsdämpfung', German Patent Application DE 35 41 201 A1, filed: Nov. 21, 1985, published: May 27, 1987.
[203] R. A. Mangiarotty, 'Control of Laminar Flow in Fluids by Means of Acoustic Energy', U.S. Patent US 4,802,642, filed: Oct. 14, 1986, patented: Feb. 7, 1989.
[204] F. Evert et al., 'Application of linear and nonlinear adaptive filters for the compensation of disturbances in the laminar boundary layer', Z. angew. Math. Mech. **80**, Suppl. 1, 85–88 (2000).
[205] M. Harper, 'Active Control of Surge in a Gas Turbine Engine', in *Proc. of the 1st Conference on Recent Advances in Active Control of Sound and Vibration*, edited by C. A. Rogers and C. R. Fuller (Blacksburg, VA, US, April 1991), pp. 133–149.
[206] W. R. Babcock and A. G. Cattaneo, 'Method and Apparatus for Generating an Acoustic Output from an Ionized Gas Stream', U.S. Patent US 3,565,209, filed: Feb. 28, 1968, patented: Feb. 23, 1971.
[207] S. Lee, 'Noise Killing System of Fans', U.S. Patent US 5,791,869 patented: Aug. 11, 1998, priority (KR): Sept. 18, 1995.
[208] Chih-Ming Ho et al., 'Active Flow Control by Micro Systems', in *IUTAM Symposium on Mechanics of Passive and Active Flow Control*, edited by G. E. A. Meier and P. R. Viswanath (Kluwer Academic Publishers, Dordrecht, NL, 1999), pp. 195–202.
[209] A. V. Boiko and V. V. Kozlov, 'Strategy of the Flow MEMS Control at Laminar-Turbulent Transition in a Boundary Layer', in *IUTAM Symposium on Mechanics of Passive and Active Flow Control*, edited by G. E. A. Meier and P. R. Viswanath (Kluwer Academic Publishers, DOrdrecht, NL, 1999), pp. 203–208.
[210] B. D. Charles, 'Blade Vortex Interaction Noise Reduction Techniques for a Rotorcraft', U.S. Patent US 5,588,800, filed: May 31, 1994, patented: Dec. 31, 1996.
[211] P. Gerontakos and T. Lee, 'Dynamic Stall Flow Control via a Trailing-Edge Flap', AIAA J. **44**, 469–480 (2006).

[212] M. L. Post and T. C. Corle, 'Separation Control Using Plasma Actuators: Dynamic Stall Vortex Control on Oscillating Airfoil', AIAA J. **44**, 3125–3135 (2006).

[213] H. T. Ngo, 'Tip Vortex Reduction System', U.S. Patent US 5,791,875, filed: Sept. 10, 1996, patented: Aug. 11, 1998.

[214] B.-H. Kim et al., 'Modeling Pulsed-Blowing Systems for Flow Control', AIAA J. **43**, 314–325 (2005).

[215] H. Preckel and D. Ronneberger, 'Dynamic Control of the Jet-Edge-Flow', in *IUTAM Symposium on Mechanics of Passive and Active Flow Control*, edited by G. E. A. Meier and P. R. Viswanath (Kluwer Academic Publishers, Dordrecht, NL, 1999), pp. 349–354.

[216] G. Wickern, 'Windkanal', German Patent Application DE 197 02 390 A1, filed: Jan. 24, 1997, published: July 30, 1998.

[217] B. Lange and D. Ronneberger, 'Control of Pipe Flow by Use of an Aeroacoustic Instability', in *IUTAM Symposium on Mechanics of Passive and Active Flow Control*, edited by G. E. A. Meier and P. R. Viswanath (Kluwer Academic Publishers, Dordrecht, NL, 1999), pp. 305–310.

[218] B. Lange and D. Ronneberger, 'Active Noise Control by Use of an Aeroacoustic Instability', Acta Acustica/Acustica **89**, 658–665 (2003).

[219] D. Guicking, 'Active Noise and Vibration Control', Annotated Reference Bibliography, 4th Edition on CD ROM (Jan. 2003), 8382 references with comfortable retrieval program GORBI (Guicking's Online Reference Bibliography). More information at http://www.physik3.gwdg.de/∼guicking and http://www.guicking.de.

[220] D. Guicking, 'Patents on Active Control of Sound and Vibration – an Overview', 2nd ed., 125-page brochure, and a comprehensive data file on CD ROM with comfortable retrieval program GOPI (Guicking's Online Patent Information). More information at http://www.physik3.gwdg.de/∼guicking and http://www.guicking.de.

Oscillations, Waves and Interactions, pp. 139–170
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-06-7

The single bubble – a hot microlaboratory

W. Lauterborn, T. Kurz, R. Geisler, D. Kröninger, and D. Schanz

Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. Experimental and numerical work on single bubbles in liquids, mostly water, is presented. The oscillation properties of acoustically driven bubbles from periodic motion to period-doubling and chaotic dynamics are reviewed. Optic cavitation as a means to prepare single bubble states in conjunction with high-speed optical observations is shown to enable detailed investigations of bubble dynamics, in particular, fast bubble collapse at various conditions. In this way, shock wave and light emission, jet and counter-jet formation near solid walls and the associated erosive action on the surface could be elucidated. Molecular dynamics studies are presented that allow a numerical view into the bubble interior that hitherto is not accessible by experiments.

1 Introduction

A single bubble in a liquid is a remarkable object. In many cases it attains a spherical shape due to surface tension and floats around in the liquid driven by the various forces that it is susceptible to, notably pressure forces of all kinds from static to acoustic, from buoyancy to drag. A single bubble, in a sense, is an artificial object, as it likes to come in clouds and swarms. However, the single bubble is the building block and starting point for describing more complex bubble configurations. Moreover, through the concentrated effort of many a scientist it has been tamed to be kept in place in a bubble trap for closer inspection [1], similar to the optical tweezer for objects susceptible to electromagnetic forces (light waves) or the ion trap for keeping ions in place with electric forces.

2 Single bubble in a sound field

It has been found that bubbles can be generated in liquids by sufficiently strong sound fields, a process called acoustic cavitation. The bubbles formed subsequently react to the pressure variations in the sound field. Sufficiently small bubbles attain a spherical shape due to the surface tension of the liquid and start to oscillate radially. Indeed, they can be considered as nonlinear oscillators driven by the varying pressure of the sound. When dispensing with other forces (as buoyancy, for instance) tractable models for the oscillation of a spherical bubble in a sound field can be formulated with varying degree of sophistication.

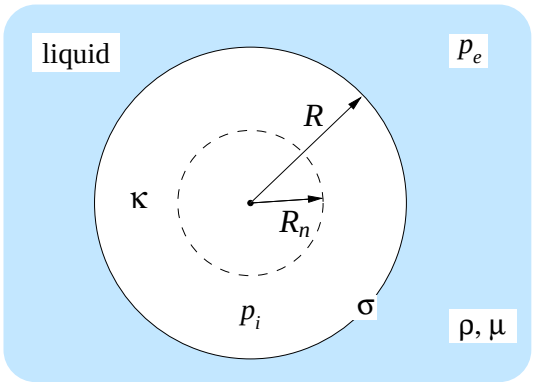


Figure 1. Descriptors for a spherical bubble: bubble radius $R(t)$, rest radius R_n, internal pressure p_i, external pressure p_e. Parameters of the liquid: density ρ, viscosity μ, and surface tension σ. The bubble medium is characterized by the polytropic exponent κ.

2.1 Bubble models

To formulate a bubble model, various parameters of the liquid surrounding the bubble and of the gas and vapour contents inside the bubble need to be specified. For a spherical bubble the dependent variables can be condensed to just one, the bubble radius R whose variation with time t has to be determined, $R(t)$. The model parameters as used here are given in Fig. 1. Besides R these are the radius of the bubble at rest, R_n, the external pressure in the liquid, p_e, and the pressure inside the bubble, p_i, further the polytropic exponent of the gas in the bubble, κ, and ρ, μ, and σ, the density, the viscosity, and the surface tension of the liquid, respectively.

The simplest model is the Rayleigh model [2]:

$$\rho R\ddot{R} + \frac{3}{2}\rho\dot{R}^2 = p_i - p_e\,, \tag{1}$$

where an overdot means differentiation with respect to time. The difference in pressure, $p_i - p_e$, drives the bubble motion. The form of the inertial terms on the left-hand side is due to the spherical three-dimensional geometry that is transformed to one radial dimension in the differential equation. Both, p_i and p_e, become functions of radius R and time t, when gas and vapour fill the bubble, and when surface tension σ, liquid viscosity μ, and a sound field are taken into account. With these inclusions the Rayleigh model takes the form [3–6]:

$$\rho R\ddot{R} + \frac{3}{2}\rho\dot{R}^2 = p_{gn}\left(\frac{R_n}{R}\right)^{3\kappa} + p_v - p_{\mathrm{stat}} - \frac{2\sigma}{R} - \frac{4\mu}{R}\dot{R} - p(t)\,, \tag{2}$$

with

$$p_{gn} = \frac{2\sigma}{R_n} + p_{\mathrm{stat}} - p_v\,, \tag{3}$$

R_n being the equilibrium radius, p_{stat} the static pressure, p_v the (constant) vapour

pressure, and

$$p(t) = -\hat{p}_a \sin \omega t \,, \tag{4}$$

the acoustic pressure, taken here to vary sinusoidally, with angular frequency ω and pressure amplitude $\hat{p}_a$. This model and some variants are called Rayleigh–Plesset models.

In this work we widely use the Gilmore model [7] that incorporates sound radiation into the liquid from the oscillating bubble, whose surface acts like the membrane of a spherical loudspeaker. It is further augmented by a van der Waals law [8] to account for a noncompressible volume of the inert gas inside the bubble. This bubble model reads:

$$\left(1 - \frac{\dot{R}}{C}\right) R\ddot{R} + \frac{3}{2}\left(1 - \frac{\dot{R}}{3C}\right)\dot{R}^2 = \left(1 + \frac{\dot{R}}{C}\right) H + \frac{\dot{R}}{C}\left(1 - \frac{\dot{R}}{C}\right) R\frac{dH}{dR} \,, \tag{5}$$

where

$$H = \int_{p|_{r\to\infty}}^{p|_{r=R}} \frac{dp(\rho)}{\rho} \,, \tag{6}$$

$$p(\rho) = A\left(\frac{\rho}{\rho_0}\right)^n - B \,, \tag{7}$$

$$p|_{r=R} = \left(p_{\mathrm{stat}} + \frac{2\sigma}{R_n}\right)\left(\frac{R_n^3 - bR_n^3}{R^3 - bR_n^3}\right)^{\kappa} - \frac{2\sigma}{R} - \frac{4\mu}{R}\dot{R} \,, \tag{8}$$

$$p|_{r\to\infty} = p_{\mathrm{stat}} + p(t) \,, \tag{9}$$

$$C = \sqrt{{c_0}^2 + (n-1)H} \,. \tag{10}$$

The additional parameters and variables in this model are the sound velocity in the liquid at normal conditions, c_0, the sound velocity at the wall of the bubble, C, the enthalpy, H, the parameters of the equation of state, where b is the van der Waals constant, and the Tait equation (7) is chosen for the liquid with its parameters A, B, and n [9].

2.2 Oscillation properties

Both models described so far consist of just one ordinary differential equation of first order, however a strongly nonlinear one. They can be viewed as describing a nonlinear oscillator, but one with really peculiar properties, for instance a varying mass during the oscillation. This leads to special oscillation features, in particular extremely strong collapses of the bubble at elevated acoustic forcing.

The equations of motion for the bubble radius as a function of time can be solved after initial conditions for the radius and for the velocity of the bubble wall have been specified. This has been done for a substantial part of the parameter space [6,10,11]. The main parameters of interest are the bubble radius at rest, R_n, and the sound field parameters pressure amplitude, $\hat{p}_a$, and sound field (circular) frequency, ω. The other parameters are connected with the properties of the liquid outside the bubble and the gas or gases and the vapour of the liquid inside the bubble.

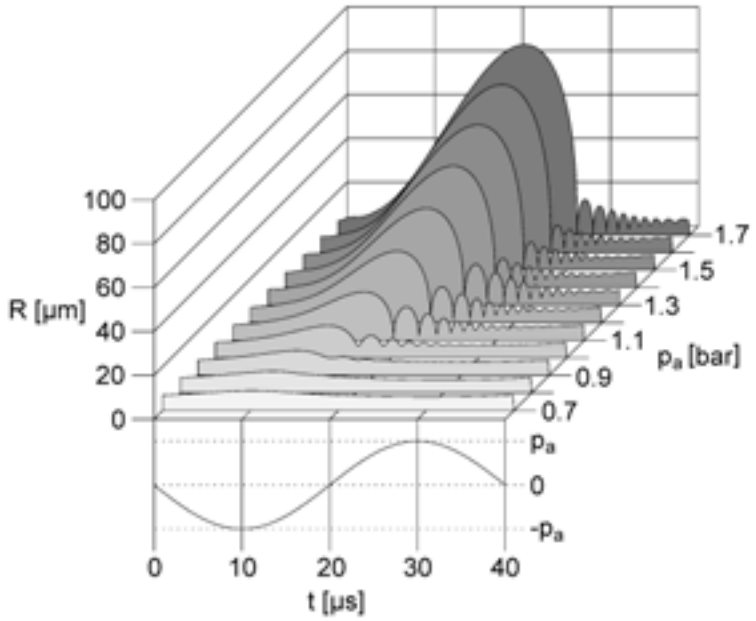


Figure 2. Synopsis of the steady-state radial oscillations of a bubble in water as a function of acoustic driving pressure. The bubble's equilibrium radius is 7 μm, the driving frequency, 25 kHz. The maximum radius strongly increases for pressures above 1 bar, which is the static pressure in the liquid.

Closed analytical solutions to the above bubble models are not known. Numerical solutions, on the other hand, are easily obtained today. When starting at some initial condition the bubble oscillator responds with transients until it settles to a steady state determined by the damping. Figure 2 gives a sequence of steady-state radius-time curves for a bubble of 7 μm radius at rest driven by a sound field of 25 kHz at various sound pressure amplitudes from 0.7 bar to 1.7 bar. The variation from an almost sinusoidal oscillation of small amplitude to one with strong elongation, fast collapse and afterbounces at compression is demonstrated.

All oscillations in the figure repeat after one oscillation of the sound field. This need not be the case [6]. The steady state oscillation may only be stably reached after two periods of the sound field, or four or eight. Indeed, the period of the steady state oscillation may go to infinity despite periodic driving of the bubble. The oscillation is then said to be chaotic, as it never repeats, albeit everything is deterministic. Acoustically driven bubbles follow a period-doubling route to chaos, where upon altering a parameter (either bubble radius at rest or sound field frequency or sound pressure amplitude) the bubble oscillation doubles its period at ever finer alteration, so that chaotic oscillations are reached at a finite parameter value. Examples of the period-doubling route to chaos are given in Lauterborn and Parlitz [12] and Parlitz et al. [10], for instance. Indications of this behaviour can already be found in an early paper by Lauterborn [6].

2.3 Response curves

Oscillators are often described by response curves. In linear systems the amplitude of the (then sinusoidal) steady state oscillation is plotted versus the driving frequency to give the amplitude response curve or resonance curve. It shows just one peak, the resonance, when the damping is sufficiently small. With nonlinear oscillators this main resonance starts to lean over to higher or lower frequencies of the driving, eventually developing hysteresis, where there exist two steady states at the same frequency of the driving each with its own basin of attraction (set of initial conditions) leading to the respective steady state (attractor). Such basins can be found in Parlitz et al. [10]. Moreover, additional resonances appear at or near rational numbers of the ratio of driving frequency and linear resonance frequency of the bubble, called harmonic resonances, subharmonics resonances and ultraharmonic (or ultrasubharmonic) resonances [6]. They also may show hysteresis. And even more, period doubling sets in along with the resonances with incomplete ("period bubbling") and complete routes to chaos.

Figure 3 shows a set of response curves with the bubble radius at rest as the control parameter instead of the driving sound field frequency. The maximum relative response, $(R_{max} - R_n)/R_n$, for a driving frequency of 20 kHz and for sound pressure amplitudes from 10 to 70 kPa is plotted versus R_n. The emerging peaks are resonances that can be labelled with two numbers, the torsion number and the period number (both natural numbers) [6,10]. Two sets of resonances are to be seen: the large peaks that fall off towards lower radii, labeled with a '1' in the denomi-

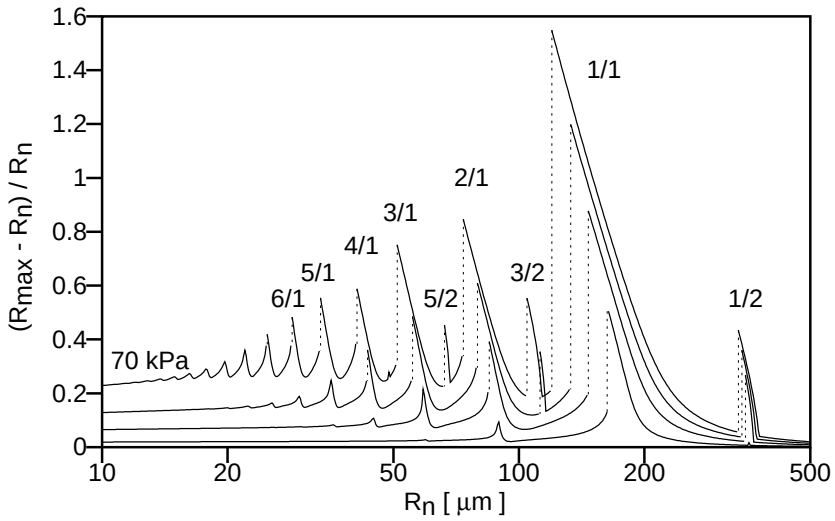


Figure 3. Response curves for bubbles of different sizes subject to driving at a fixed frequency of 20 kHz. Driving amplitudes are 10 (bottom curve), 30, 50, and 70 kPa (upper curve). Initial condition is the respective bubble at rest. The logarithmic radius scale stretches the harmonic resonances for better viewing.

nator, and the smaller ones in between, labeled with a '2' in the denominator. The first set constitutes the main resonance (resonance of order 1/1) and the harmonic resonances of order $n/1$ with $n = 2, 3, \ldots$, given by the period number 1 and the torsion numbers $n > 1$. The second set belongs to the set of subharmonic resonances, starting with the main subharmonic resonance, 1/2, to the right of the main resonance, and going on with the 3/2, 5/2, 7/2 resonances (also called ultrasubharmonic or ultraharmonic resonances) to the left of the main resonance. The appearance of the nonlinear resonances follows a Farey tree [10]. For proper visualization of the sequence of resonances that cluster towards smaller bubble radii the R_n scale is logarithmic.

There appear straight vertical dotted lines in the diagram. At these radii R_n the steady state oscillation behaves nonmonotonously with the bubble radius at rest. The reason is the overturning of the resonances leading to a small or a large amplitude oscillation depending on the initial condition chosen. In the language of nonlinear dynamical systems theory [12] there are two coexisting attractors. They span a region of hysteresis, a region where two different oscillations are stable. Each of the attractors then has a set of initial conditions from where it is approached: its basin of attraction. The diagram was calculated with the respective bubble starting from its rest position ($R = R_n, \dot{R} = 0$). Then just one of the two attractors will be reached and the jump occurs when the boundary between the two basins of attraction sweeps over the rest position upon varying the parameter, here R_n.

When driving the bubbles stronger, the resonances as shown in Fig. 3 fill with period-doubling cascades to chaos and sudden onset of chaotic oscillations (Fig. 4). Also a giant response occurs for smaller bubble radii. Surface tension has a strong

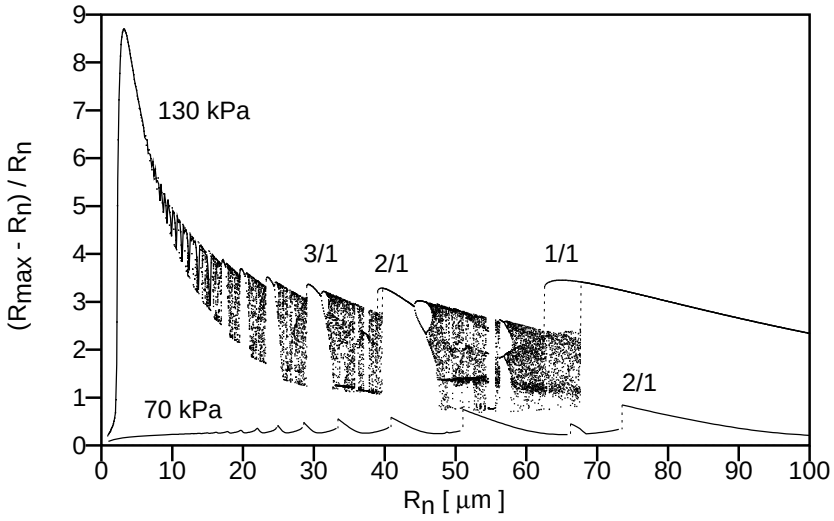


Figure 4. Response 'curve' for bubbles of different sizes driven at a fixed frequency of 20 kHz with $\hat{p}_a = 130$ kPa. For comparison the response for $\hat{p}_a = 70$ kPa is also given. The linear radius scale is chosen for properly demonstrating the giant response near $R_n \approx 5\,\mu$m.

influence in this region through forcing the sudden, but smooth, decay that occurs at lower bubble radii. Without surface tension this decay would be missing and the response would level off at a high expansion oscillation. To put the giant response at small bubble radii into perspective, Fig. 4 gives the response for $\hat{p}_a = 130\,\mathrm{kPa}$ and 70 kPa at 20 kHz for bubbles from 1 μm to 100 μm on a linear radius scale. The by far greater amplitude of the giant response with respect to the main resonance is easily noticed. The relation gets even more pronounced at higher driving amplitudes. However, then the radial oscillation stability may be lost as is inevitable at sufficiently strong driving.

2.4 Parameter space diagrams

The response curves already condense the information on the oscillation properties of a bubble as not the full oscillation is retained but only the maximum elongation, or a set of points in the case of a chaotic oscillation. An even more condensed survey of information on the response of a bubble to a sound field can be given in the form of a parameter space diagram, often also called phase diagram (Fig. 5). There, over a space of parameters, best only two parameter variables, just the type of the resulting oscillation is plotted into the respective volume (area). The areas are determined by the boundaries of the bifurcation sets of the bubble oscillators. The type of oscillation inside these boundaries may be colour coded for better vision or elsewise made discernable. Oscillation types are, for instance, discerned by the period of the resulting oscillation, that is period 1 or period 2, or generally period n, where n can be any natural number up to infinity. Also, chaotic oscillations may be classified

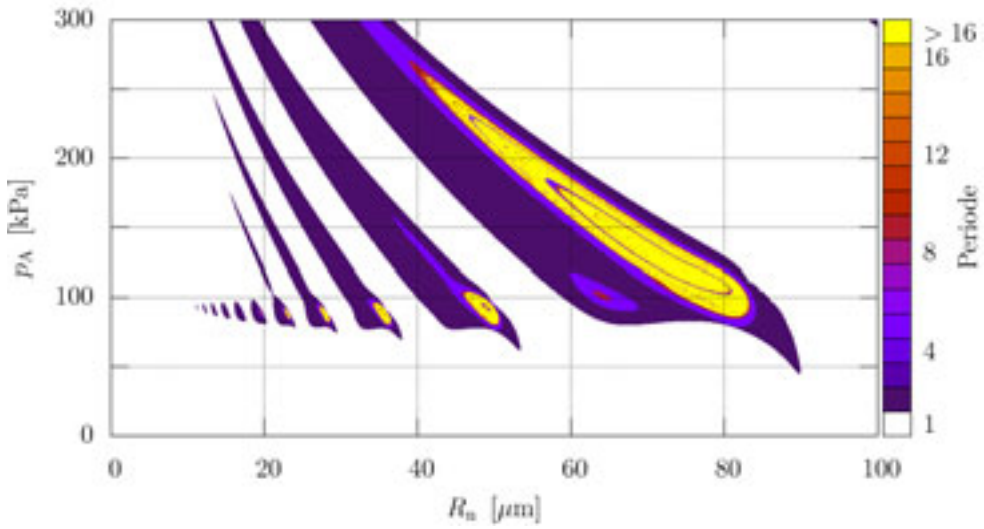


Figure 5. Parameter space diagram obtained with the Keller-Miksis bubble model depicting the periodicity properties of bubble oscillations in the $(R_n, \hat{p}_a)$-parameter plane (calculation by P. Koch).

by types according to from which resonance they sprang off. Phase diagrams for bubble oscillators can be found in Parlitz et al. [10]. Figure 5 presents a parameter space diagram calculated with the Keller-Miksis model [13] in the $(R_n, \hat{p}_a)$-parameter space. The harmonic resonances shown form building blocks of period prolongation that repeat with each harmonic resonance.

2.5 Single bubble in an acoustic trap

A single bubble can be trapped in the pressure antinode of a standing sound field [1]. This facilitates experiments with single bubbles considerably. The bubble oscillations [14], the shock wave emission [15] and the light emission [16] have been studied thereby in great detail. The simplest arrangement consists of a rectangular glass cuvette with just one piezoelectric transducer glued to the bottom as shown in Fig. 6. The thin wire sticking into the cuvette from the upper left side is of platinum and serves as bubble generator via a current pulse for introducing a bubble to be trapped into the liquid. A stably trapped bubble is to be seen by its bluish light emission. As the bubble oscillation repeats itself with great precision each acoustic cycle, whereby the bubble also keeps its position, the bubble can be photographed with high resolution in time and space by use of a long distance microscope. Figure 7 shows one cycle of a bubble oscillation driven at 24.1 kHz and taken in backlight conditions. The time difference from one frame to the next is 500 ns. The spherical bubble shows

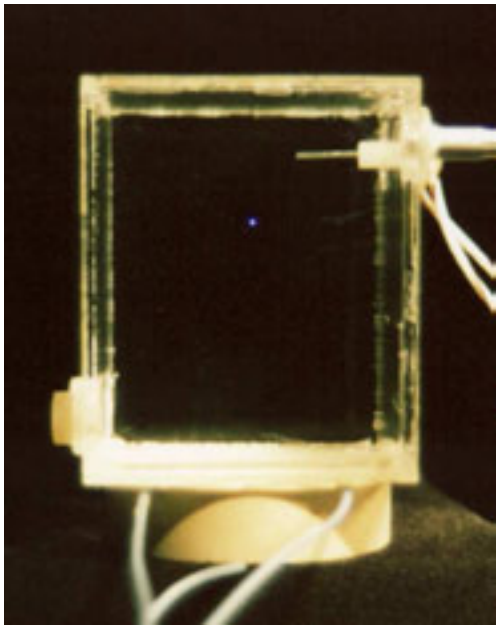


Figure 6. Photograph of a cubical bubble trap with the transducer glued to the bottom. The bluish point in the middle of the upper part of the cuvette is the light emitted by a stably oscillating bubble. Exposure time 20 min.

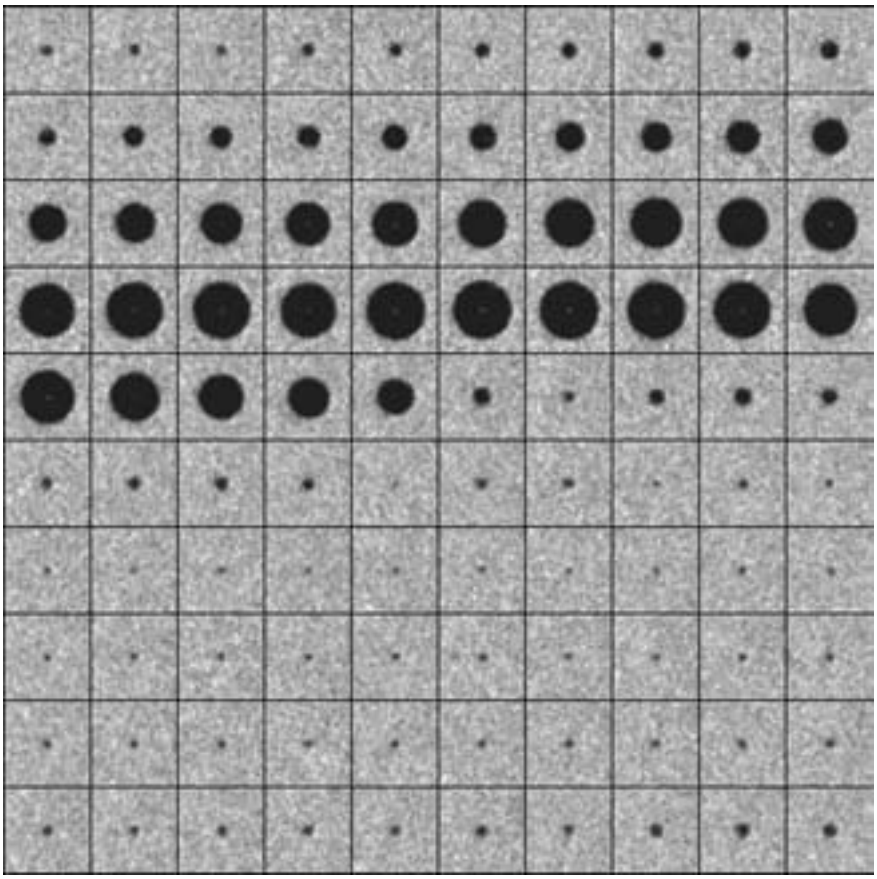


Figure 7. Photographic series of a trapped sonoluminescing bubble driven at 21.4 kHz. Time between frames: 500 ns, frame size: 160 μm by 160 μm. As the bubble oscillates very stably, the images were taken one by one (exposure time 5 ns) with increasing delay and assembled into the series lateron.

black on a brighter background as the illuminating light is deflected off the bubble surface. A slow growth of the bubble and a fast collapse with several afterbounces are observed. From the photographic images the varying radius of the bubble can be measured and compared with theoretical models of bubble oscillation. Figure 8 shows a comparison for a bubble trapped in a water-glycerine mixture driven at 21.4 kHz and 132 kPa (measured values). The data input to the Gilmore model are: ambient pressure p_0 =100 kPa, vapour pressure p_v =0 kPa, equilibrium radius R_n =8.1 μm, density of the liquid ρ =1000 kg/m^3, viscosity μ =0.0018 Ns/m^2 (measured) and surface tension σ =0.0725 N/m. The gas within the bubble is assumed to obey the adiabatic equation of state for an ideal gas with κ =1.2. A quite good fit could be achieved even for the first three afterbounces, except for the strong collapse where data are missing.

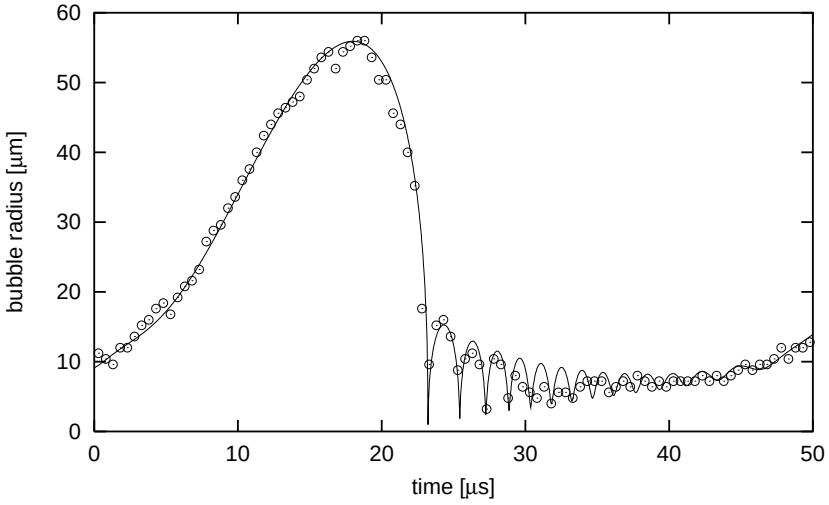


Figure 8. Comparison of the experimental single bubble oscillation (open circles) as shown in Fig. 7 with numerical simulations based on the Gilmore model (solid line).

3 Single light-induced bubbles

The single bubbles in an acoustic trap must adjust to special conditions to be trapped. Thus their size goes with the driving frequency, being larger for lower frequency, and the gas content of the liquid should be low for best stability of the oscillation. Also, in the course of oscillation the bubble contents alters in adjustment to the outer and self-produced conditions via diffusion of mass and heat and via chemical reactions triggered by high temperatures in the inside during collapse [17,18].

A larger variety of single bubbles can be produced by concentrating light energy in the focus of a laser beam forcing dielectric breakdown of the liquid with concomitant bubble formation [19]. In analogy to acoustic cavitation this bubble forming process has been called optic cavitation. As the location of the bubble and its instant of generation are precisely known, similarly sophisticated measurements with high-speed optical equipment are possible on bubble oscillation, shock wave radiation and light emission as in the case of acoustically trapped bubbles. Moreover, additional measurements are possible on the special dynamics of bubbles near walls.

Bubble dynamics near boundaries or obstacles is outside the scope of trapped bubbles, as bubble stability cannot be maintained. It is, however, of strong relevance to hydrodynamic cavitation with its erosion problem and also to ultrasonic cleaning. Even more so, bubble interaction can be studied with light-induced bubbles by producing two or more bubbles simultaneously at preselected places and of preselected sizes or with a time shift between them. Some of the possibilities have already been explored [20–24], the majority of cases, however, waits to be investigated.

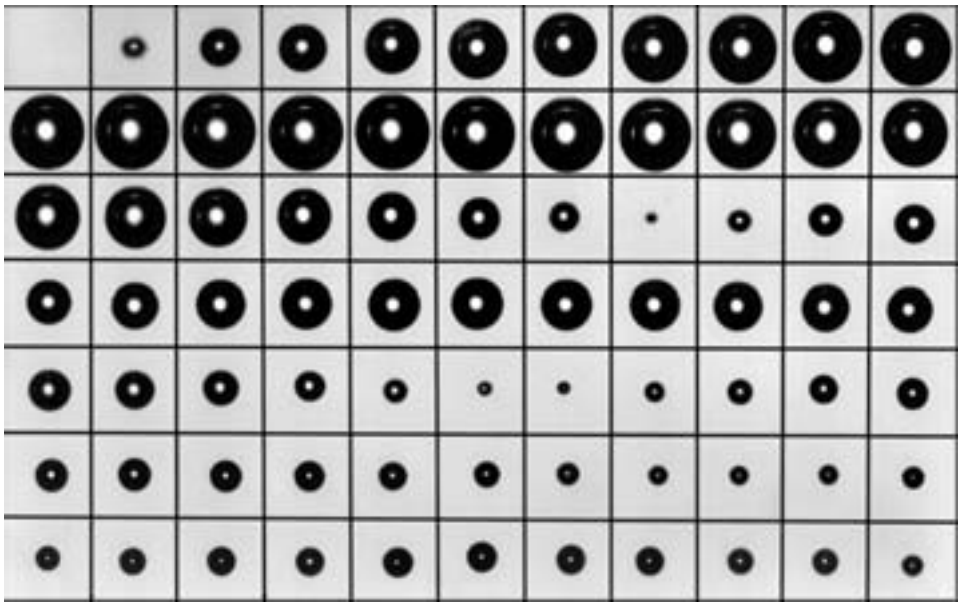


Figure 9. Dynamics of a laser-produced spherical bubble in silicone oil of viscosity 4.85 poise observed at 75 000 frames per second. Maximum bubble radius is about 2 mm.

3.1 Laser-bubble dynamics

The following figures present some examples of single bubble dynamics under various circumstances. Figure 9 shows the free oscillations of a single spherical bubble produced in silicone oil attaining a maximum bubble radius of about 2 mm, substantially larger than the few 10 μm of acoustically trapped bubbles. The oscillation is damped by the viscosity of the liquid and the bubble attains its equilibrium radius after about four oscillations. The bright spot in the middle of the bubble is a result of the backlight. In a nicely spherical bubble of this size the light traverses the bubble undeflected and nearly undeflected along and near perpendicular incidence of the illuminating light on the bubble surface.

Figure 10 shows the collapse of a single bubble near a solid wall. The sequence taken at 300 000 frames per second starts near the first collapse and shows the characteristic protrusion of the bubble wall towards the surface. The liquid jet is hidden inside this protrusion and is best visible in the rebound phase as a dark line inside the bright spot at the center of the cavity where the backlight can pass undisturbed through the smooth bubble surface. After some time the long tube of gas and vapour with the jet inside gets unstable and decays into several tiny bubbles. The surface of the bubble thereby reapproaches its spherical shape.

The formation and strength of a jet inside the bubble depends on the distance s of the bubble center from the wall. To be more exact, the normalized distance or stand-off parameter

$$\gamma = \frac{s}{R_{\max}} ,$$

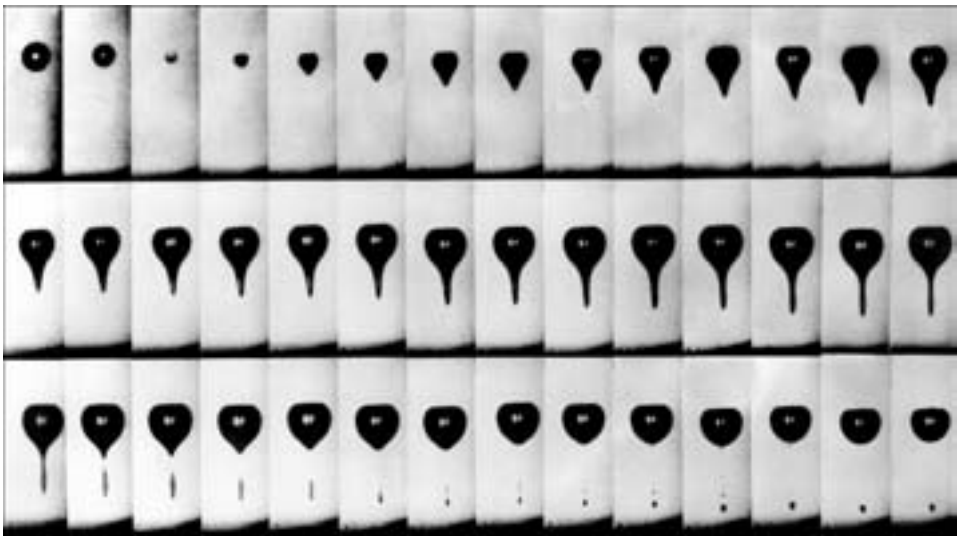


Figure 10. Dynamics of a laser-produced spherical bubble in water of reduced surface tension in the neighbourhood of a plane solid boundary observed at 300000 frames per second. Size of the frames is 6.7 mm × 2.7 mm.

where $R_{\max}$ ist the maximum radius of the bubble and s the distance from the wall at that instant, has turned out to be a good measure of perturbation of the bubble collapse by a neighbouring wall.

The development of a jet by involution of the part of the bubble opposite to the solid wall, the formation of a toroidal bubble or bubble vortex ring, and the rebound of the torus bubble with the formation of a 'counterjet' is exemplified by Fig. 11. The photographic series is taken at one million frames per second for better resolving the dynamics and at an angle of 45° from above the wall with front illumination. The normalized distance to the wall has the quite large value of γ =2.6. The jet is broad and after penetrating the opposite bubble wall turns the bubble into a short living torus bubble or vortex ring, the life time being less than a microsecond. Also within less than a microsecond the torus bubble has rebounded and is expanding whereby a 'counterjet' develops sticking out upwards perpendicular to the torus ring and with obviously little physical connection to the expanding torus bubble. It is produced by the tension part of the torus shock wave developing upon collapse of the torus as secondary cavitation [23]. It consists of many tiny bubbles that nucleate in the strong negative pressure region along the perpendicular to the torus ring by confluence of the shock/tension waves emanating from the torus ring. Figure 12 shows the evolution of this secondary cavitation appearing on top of the bubble (torus) photographed in side view for a bubble of $R_{\max} = 1.5$ mm and $\gamma = 1.4$.

When femtosecond laser pulses are used for bubble generation the high intensity in the laser beam alters the properties of the medium in which it propagates, in particular its index of refraction. In water this leads to self-focussing of the beam and self-guiding or channelling. This effect may be used to produce elongated bubbles for

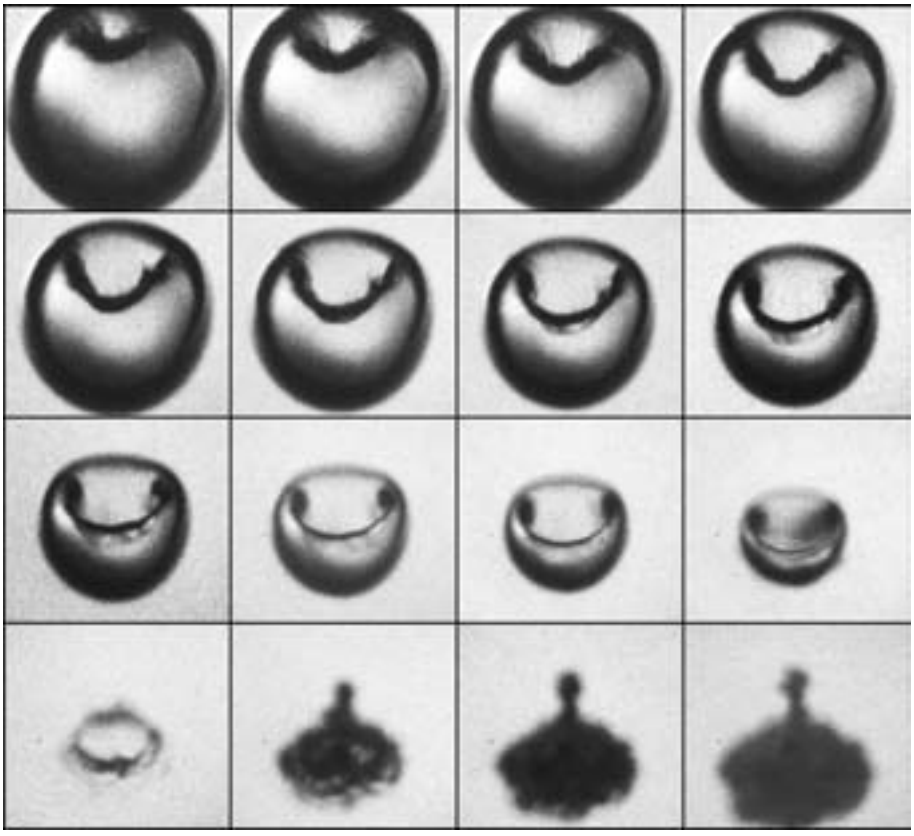


Figure 11. Bubble collapse as photographed from an angle of 45° from above the solid wall with front illumination. Maximum bubble size is $R_{\max}$ =1.5 mm, normalized bubble distance γ is 2.6, the frame size is 1.1 mm (height) by 1.2 mm (width) (from Ref. [23]).

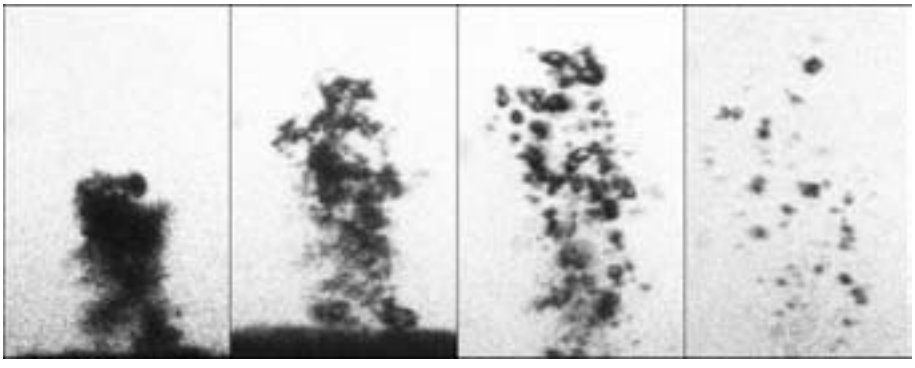


Figure 12. Evolution of the secondary cavitation event ('counterjet') for a bubble of $R_{\max} = 1.5$ mm and $\gamma = 1.4$. Four frames taken at 5, 50, 100 and 200 μs after collapse are shown. Frame size is 1.21 mm by 0.77 mm (from Ref. [23]).

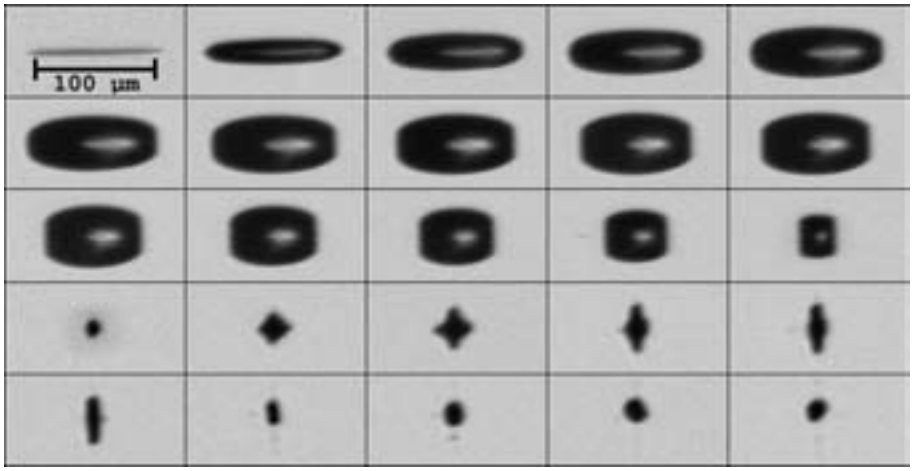


Figure 13. Shape dynamics of an elongated femtosecond-laser-induced bubble. Interframe time 400 ns; laser energy 0.8 μJ per pulse.

studying their dynamics. Figure 13 gives an example of an initially stretched bubble of about 100 μm length generated by a femtosecond laser pulse of 0.8 μJ energy. The photographic series depicts the typical scenario. The length of the bubble does not increase, instead the bubble heads for a more spherical shape. The elongated ends, however, involute early in the dynamics leading to a quadratic, after rebound a starlike and later flat appearance unless the oscillation settles to a near spherical shape, all in about 12 μs only.

In Fig. 14 the dynamics of an elongated bubble as produced by a femtosecond light pulse is compared with numerical calculations done with the boundary integral method [25]. The flattening and involution of the bubble surface from both ends, as expected from the experiments, is reproduced in the calculations. Although a perfect fit cannot be expected as the initial conditions cannot be specified to high accuracy, the overall agreement in the dynamics is remarkable.

3.2 Laser-bubble shock waves

Light-induced bubbles come with a shock wave upon generation as the material of the fluid is strongly heated and develops a large pressure that is radiated into the liquid. The bubble expansion follows on a slower time scale. When the bubble collapses an intense shock wave is radiated in a kind of inverse process where the bubble contents are again compressed strongly, simultaneously with the surrounding liquid near the surface of the bubble. This again leads to the emission of a shock wave. Also the subsequent collapses of the bubble emit shock waves, when the viscous damping of the liquid is low, as for water. These shock waves can be measured with sufficiently fast hydrophones (Fig. 15). A glass fibre is brought near to the breakdown and collapse point to capture the shock waves via the change of the reflection properties of light at the end of the fibre by the density alteration of the liquid upon compression going with

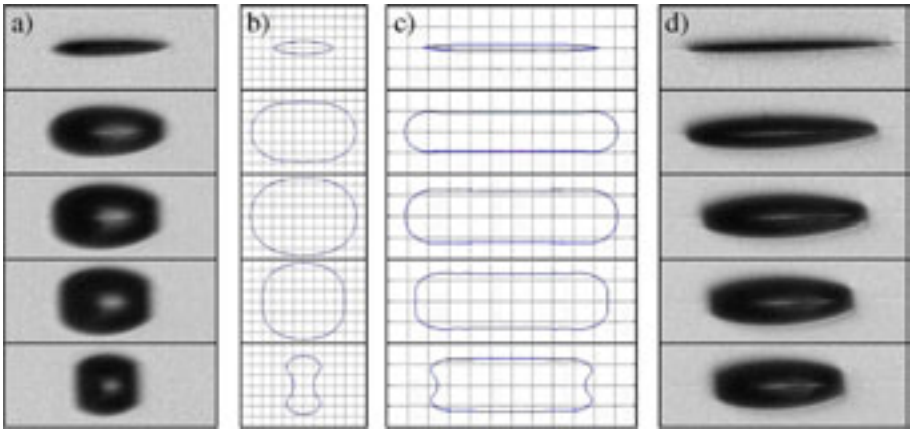


Figure 14. Shape dynamics of two elongated femtosecond-laser-induced bubbles. The height of the frames is 47 μm. Comparison with theory courtesy of N. Pelekasis and K. Tsiglifis [25].

the shock wave. Figure 16 gives the result of a shock wave pressure measurement series with different bubble sizes R_{max}. The upper and lower boundaries of the shaded area give the error bounds of the measurements. The pressures obtained at distance d from the breakdown site have been extrapolated to the minimum bubble radius by using a $1/r$-dependence for the amplitude, where the minimum radius has been obtained by a numerical fit to the bubble collapse. About 10 kbar are reached for a bubble collapsing from a radius of 500 μm and about 25 kbar when collapsing from a radius of 3 mm. Unfortunately the collapse pressure from trapped bubbles cannot be measured this way as these bubbles are extremely sensitive to a disturbance nearby and also of lower shock wave strength, because they reach a lower maximum size of less than about a 100 μm and thus have less total energy. But extrapolating the pressure values down to smaller radii R_{max} gives strong evidence of collapse pressures in the range of a few 1000 bar for standard sonoluminescing trapped bubbles.

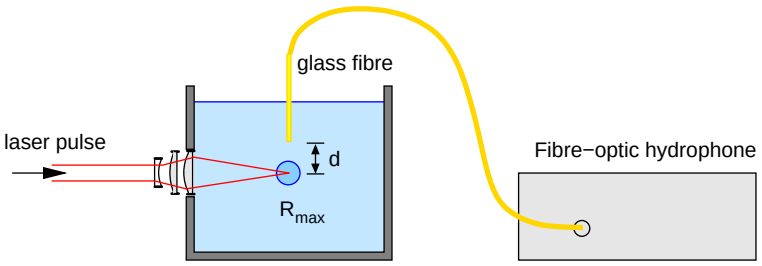


Figure 15. Experimental arrangement with a fibre-optic hydrophone for the measurement of shock waves from laser-induced breakdown and of bubble collapse shock waves. The distance d between the fibre tip and the bubble can be varied.

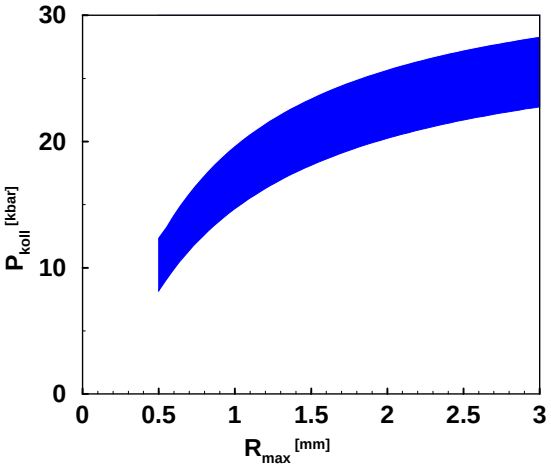


Figure 16. Collapse pressure at minimum bubble radius from a laser-induced bubble versus maximum attained radius. The collapse pressures are extrapolated from pressure-distance measurements taken at the corresponding maximum bubble radii, $R_{\max}$. The shaded region depicts the error margins of the extrapolation.

3.3 Laser-bubble erosion

It is long known that cavitation bubbles can damage solid surfaces. Mainly two characteristic effects are believed to be responsible for the destructive action of cavitation bubbles: the emission of shock waves upon collapse of the bubble and the generation of a high-speed liquid jet directed towards the solid boundary. The main results obtained so far are the following [22]. Damage is observed when the bubble is generated at a distance less than twice its maximum radius from a solid boundary ($\gamma \leq 2$). The impact of the jet contributes to the damage only at small normalized distances ($\gamma \leq 0.7$). The largest erosive force is caused by the collapse of a bubble in direct contact with the boundary, where pressures of up to several GPa act on the material surface. Bubbles in the ranges $\gamma \leq 0.3$ and $\gamma = 1.2$ to 1.4 cause the largest damage.

Figure 17 gives an example from the numerous measurements done. It shows the collapse of a bubble immediately touching the surface (γ =0.5) with high time

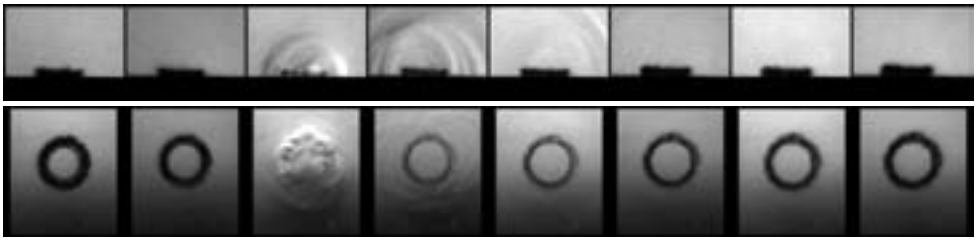


Figure 17. First collapse of a laser-induced bubble near a solid wall with $\gamma = 0.5$ taken at one million frames per second. The maximum radius of the bubble was 1.45 mm. Upper sequence: side view; lower sequence: bottom view trough the transparent solid wall. Both views together demonstrate the transformation of the spherical bubble into a toroidal bubble after the jet has formed.

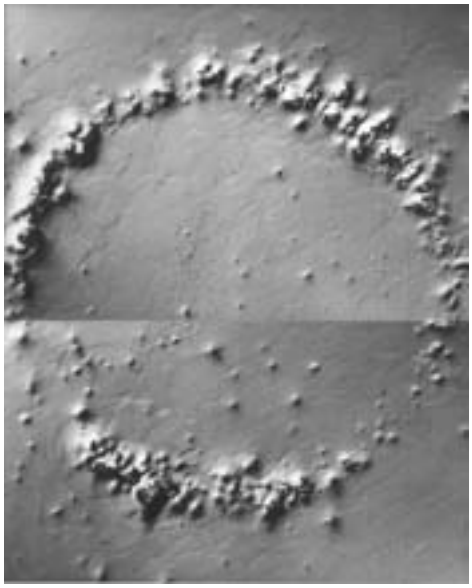


Figure 18. Damage on an aluminium surface generated by 100 laser-induced bubbles. The maximum radius of the bubbles was 2.00 mm, the stand-off parameter, $\gamma = 1.4$, the frame width, 2.8 mm.

resolution of one million frames per second. Both side and bottom view through the transparent solid wall are presented. The jet through the bubble has already touched the surface (inner bright spherical disk) leaving a toroidal bubble that spreads in size on the surface. The torus bubble ring collapses with the emission of many distinct shock waves. This indicates that the torus gets unstable upon collapse leading to several collapse sites around the torus. The shock waves seem to have been emitted directly from the surface of the wall.

The collapses along the torus bubble lead to distinct erosion pits on the solid surface. This is demonstrated in Fig. 18, where 100 laser-induced bubbles of the same size of R_{max} =2 mm were applied at the same location with a stand-off parameter of γ =1.4. The remarkable result is that, although the bubble at maximum elongation does not touch the surface and a strong jet is produced upon first collapse (as known from corresponding experiments) the damage accumulates along a ring, the location of the torus bubble to which the spherical bubble transmutes during its collapse and rebound.

3.4 Laser-bubble luminescence

Strongly collapsing bubbles emit light in short pulses at maximum compression. This is not only true for acoustically excited bubbles in sonoluminescence, but also for laser-generated or shock-excited bubbles. Therefore, the phenomenon may appropriately be called cavitation luminescence. With single laser-induced bubbles the luminescence can be studied in various situations, and its dependence on bubble size, external pressure and bubble sphericity, for instance, can be investigated. Furthermore, specific differences between single- and multi-bubble sonoluminescence can be studied with optically produced few-bubble configurations.

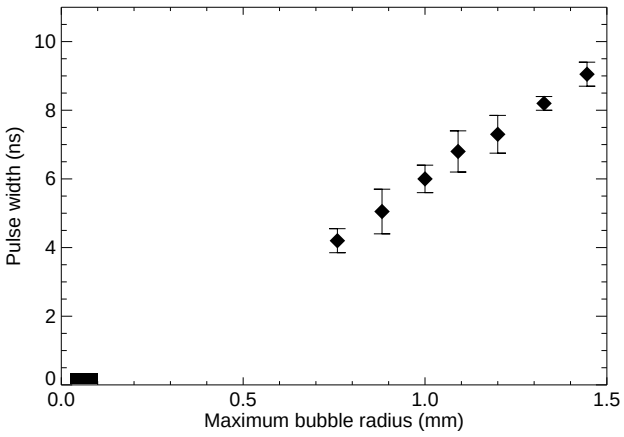


Figure 19. Pulse width of the light emission upon collapse of a laser-induced bubble versus the maximum attained radius.

Figure 19 shows the luminescence pulse width of laser-generated, freely collapsing bubbles in dependence on the maximum bubble radius from where the collapse starts. The pulse width increases linearly with bubble size and is quite long for millimetre-sized bubbles (in the nanosecond range, and thus accessible to high-speed photodetectors) as compared to the small trapped bubbles collapsing from maximum sizes of about 100 μm. The little box near the origin indicates the pulse widths measured for small trapped bubbles [16].

Current theories of single-bubble sonoluminescence attribute the light emission to a strong heating of the bubble medium by converging compression or shock waves launched within the bubble shortly before the collapse is stalled [17]. The energy concentration by this mechanism should therefore depend critically on the degree of sphericity imparted to these waves. Experimental evidence for this relation is given by Fig. 20 which shows how the light diminishes with increasing perturbation of the bubble collapse [26]. The asphericity acquired during collapse is controlled by a plane solid surface whose distance to the breakdown site is altered. A bubble far from the boundary collapses spherically (large γ), a bubble nearer to the boundary collapses the more aspherically the nearer it is to the boundary. An astounding result was found, namely that the light emission ceases at the quite large value of $\gamma = 4$. This means that the bubble must retain a sufficiently round shape upon collapse in order to produce luminescence light. Further experimental results corroborating this hypothesis are reported in the next section.

4 Laser-induced bubbles in a sound field

In order to keep a bubble stably oscillating in an acoustic bubble trap special conditions have to be met. The bubble's free radial motion may not be disturbed, so it has to be kept well away from boundaries and other bubbles. The net average force

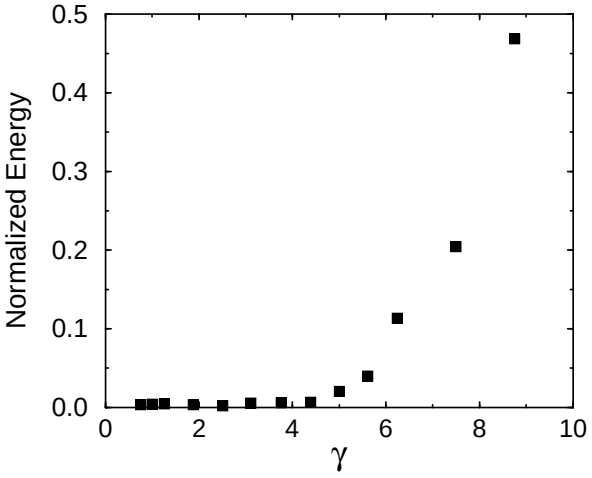


Figure 20. Luminescence of laser-induced bubbles collapsing asymmetrically near a solid wall. Dependence of the light energy normalized to the light emitted from a spherical bubble on the stand-off parameter γ.

due to pressure gradients in the liquid must vanish not to induce bubble translation. Also, streaming of the liquid must be avoided for the same reason. The amplitude of the sound field is constrained to a small interval where the bubble is in a stable diffusive equilibrium, in accord with the liquid's gas content. The bubble collects a certain amount of noble gas and has a rest radius of a few micrometres, typically. Thus by means of stable bubble trapping a limited range of the bubble's parameter space is available for exploration. The advantage of the method is that the bubble can be scrutinized easily, e. g. by high-speed photography, because it stays at a fixed location and can oscillate clock-like for many hours.

By means of optic cavitation steady and transient bubble dynamics in a sound field can be explored in a much larger parameter space than available with bubble trapping, while the advantage of being able to image the bubble dynamics at least for a limited amount of time is preserved. Furthermore, bubble translation and the acoustic and hydrodynamic forces on a bubble, e. g. drag forces, as well as the interaction of two or more bubbles can be studied in a well-defined setting. Thus the combination of optic generation and acoustic driving represents an ideal laboratory to study many aspects of single bubble behaviour and also to control bubble motion by carefully chosen parameter changes – an aspect that up to now has not been pursued extensively in experiments.

Fig. 21 shows the basic experimental setup for optic cavitation in a sound field. As described previously the bubble is generated by focussing a (nanosecond to femtosecond) laser pulse in the liquid. Simultaneously an ultrasonic field is applied in the cuvette at one of its main resonance frequencies, preferably the (1,1,1) mode, to achieve high sound pressures. An essential experimental feat is that the laser pulse can be applied at a well defined phase of the acoustic field which yields reproducible initial conditions for the ensuing bubble dynamics. For that purpose electronic devices were built that synchronize the laser operation with the acoustic field.

One issue of central importance in bubble physics is the question of how large a degree of energy focussing can be achieved in the violent collapse of a cavitation bub-

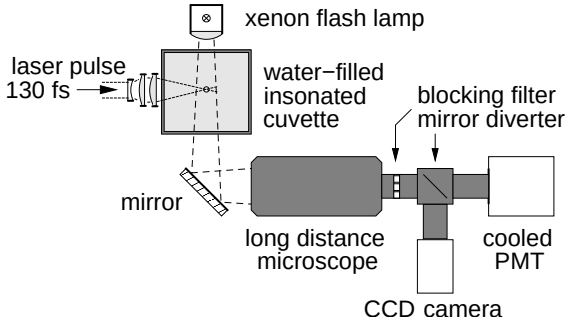
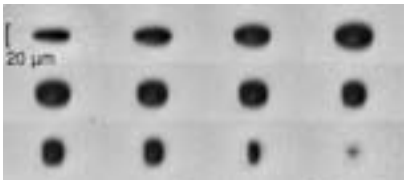


Figure 21. Experimental setup for studying the dynamics of laser-generated bubbles in an ultrasonic field.

ble. Even at the rather moderate acoustic driving pressures (∼1.2 to 1.4 bar) used in stable single bubble sonoluminescence experiments a high degree of energy focussing is obtained, leading to temperatures at the bubble center in excess of 10000 K. By employing optic cavitation in a sound field, much higher acoustic pressures up to the cavitation threshold in the liquid can be utilized for driving the bubble.

Then, under suitable conditions, the bubble's energy available for collapse is mainly supplied by the sound field and not by the laser pulse, and is only limited by the maximum radius the bubble can grow to in the negative-pressure phase of the acoustic oscillation without getting unstable. As an example, in Fig. 22 the dynamics of a femtosecond-laser-generated bubble in water without acoustic excitation (top left) is compared with the case where a bubble generated with the same laser pulse parameters is exposed to a sound field of moderate amplitude (p_a =159 kPa, right).



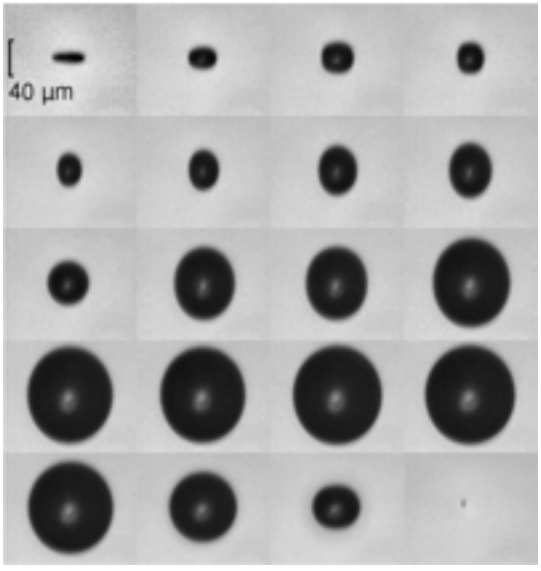


Figure 22. Top: fs-laser-generated bubble in water without sound field. Frame separation: 200 ns, collapse time ≈ 2.2 μs, maximum bubble size ≈ 30 μm×55 μm. The bubble is elongated due to self-focussing of the laser pulse. Right: bubble generated at the same conditions in a sound field of p_a ≈ 159 kPa and f_a ≈ 44 kHz. Frame separation: 800 ns. Maximum bubble diameter ≈ 100 μm, collapse time ≈ 15 μs.

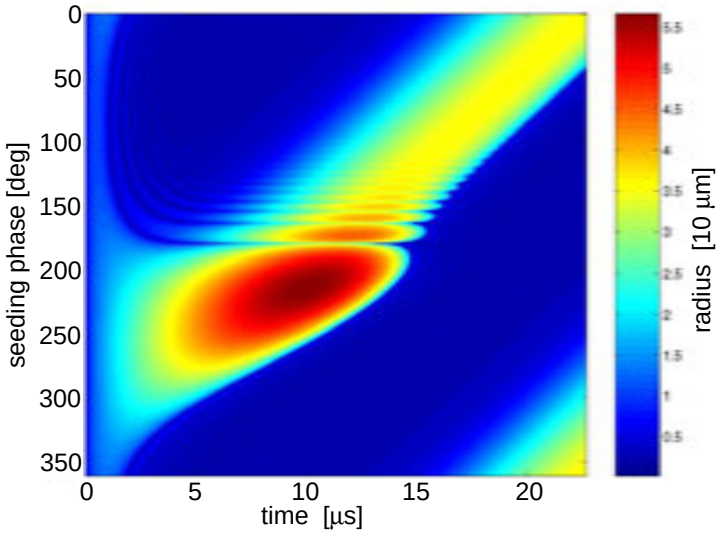


Figure 23. Numerical simulation of the transient dynamics of a laser generated bubble as a function of seeding phase and time. The bubble radius has been calculated by the Gilmore model and is shown colour-coded, the initial conditions and sound field parameters are taken from the experiment as described in the text. For the phase interval where luminescence of transient bubbles can be observed in the experiment this calculation gives a peak radius that is larger than that of the steady oscillation.

Because of the small laser pulse energy used in this experiment the non-driven bubble remains small and collapses shortly after its inception. Due to nonlinear optical effects it attains a cylindrical shape (see Sect. 3.1). Such a bubble does not emit any detectable luminescence light. The driven bubble is generated at the beginning of the negative-pressure phase. In the presence of sound the bubble grows significantly larger and becomes more spherical. The maximum radius that can be obtained for a given driving level depends, of course, on the instant of bubble inception relative to the pressure cycle, the seeding phase φ_s [1]. Figure 23 shows a numerical simulation of the first transient oscillation cycle of a laser-generated bubble as a function of seeding phase based on the Gilmore model. The bubble's initial conditions have been taken from the corresponding experiment and are the same for all phase values [27]. Around $\varphi_s \approx 220°$ the simulation yields a particularly large maximum radius as the bubble is optimally expanded by the sound field and the laser pulse energy is utilized as well.

For seeding phases of $\varphi_s < 180$ the bubble behaviour is different. Because of the overpressure now acting on the bubble just generated, it collapses quickly and stays small until the sound field creates tension to expand it again. For $\varphi \to 0$

[1] The seeding phase is defined to be zero at the beginning of the positive-pressure phase of the sound field.

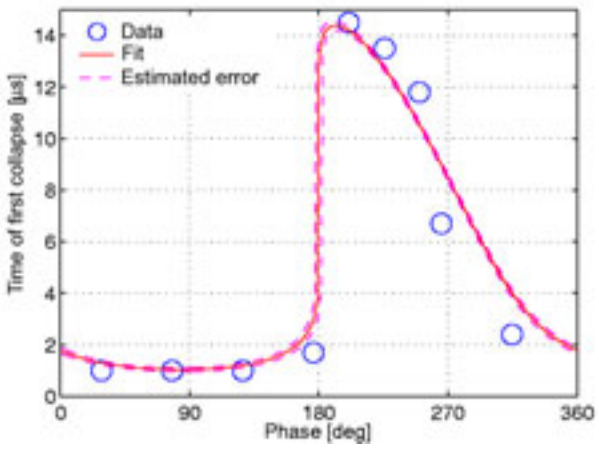


Figure 24. Experimentally measured times of first collapse of the optically generated bubble (circles) and numerically fitted curve obtained with the Gilmore model. The fit yields the phase calibration and the acoustic amplitude that cannot be measured by a hydrophone in situ without significantly disturbing the sound field.

a large maximum radius is attained later and later in the acoustic cycle (Fig. 23). This leads to a discontinuity in collapse times vs. seeding phase as shown in Fig. 24. Experimental values are plotted along with numerical data obtained with the Gilmore model for parameters that are fitted to yield minimum deviation. By comparing experimental data with numerical results from a proven model the bubble can be used as a pressure sensor in a self-consistent way: the fit yields the zero of the seeding phase and the amplitude of the acoustic pressure at the bubble position. The jump in collapse time corresponds to the point where the sound field is able to reverse the bubble motion before the first, immediate collapse occurs. This method of pressure determination has the advantage that no hydrophone, which invariably tends to disturb the sound field, has to be immersed in the liquid, and that it can also be used with a closed cuvette.

Femtosecond laser pulses have been used for bubble generation in these experiments to reduce the amount of non-condensable gas that is created in the breakdown process. Femtosecond pulses have a low breakdown threshold and thus a small plasma volume, so the gas production is minimized. Less gas means a more violent collapse as the bubble can shrink to a smaller volume before the pressure in the bubble medium counteracts the inertially dominated liquid inflow. In this way a high energy concentration and collapse strength may be achieved, that can be assessed by measuring bubble luminescence and shock wave emission as its indicators, and their dependence on acoustic driving pressure, for example.

Figure 25 presents an investigation of bubble luminescence for optically generated bubbles in an ultrasonic field. Light emission of the bubbles in dependence on their seeding phase φ_s was measured for bubbles of the type shown in Fig. 22(right). While,

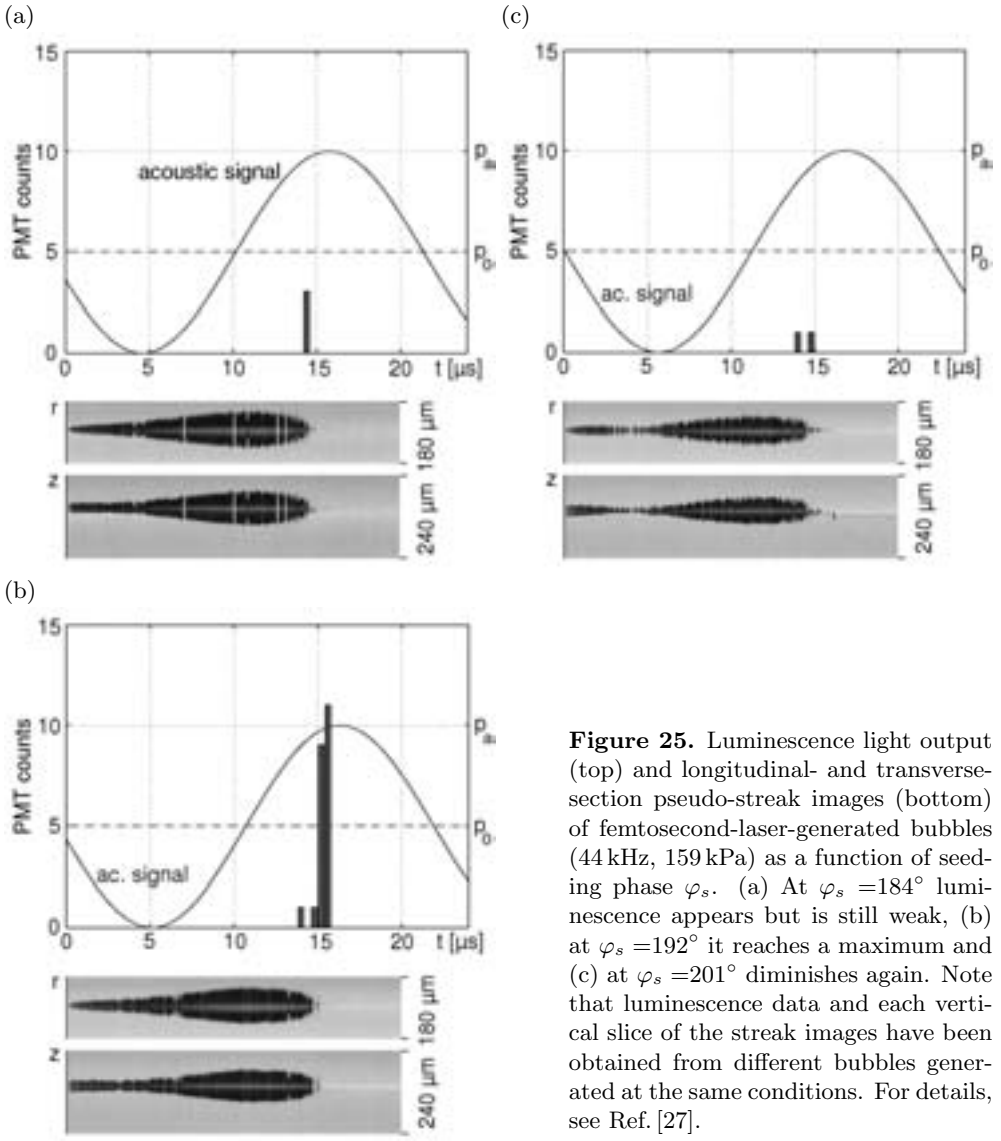


Figure 25. Luminescence light output (top) and longitudinal- and transverse-section pseudo-streak images (bottom) of femtosecond-laser-generated bubbles (44 kHz, 159 kPa) as a function of seeding phase φ_s. (a) At φ_s =184° luminescence appears but is still weak, (b) at φ_s =192° it reaches a maximum and (c) at φ_s =201° diminishes again. Note that luminescence data and each vertical slice of the streak images have been obtained from different bubbles generated at the same conditions. For details, see Ref. [27].

as mentioned before, the non-driven bubbles do not emit light, weak luminescence of the driven bubbles could be detected in two small intervals of the seeding phase. The first interval around $\varphi_s = 192°$ is shown in the figure and nearly coincides with the region of maximum bubble expansion seen in Fig. 23. Closer analysis of the bubble shape reveals that light emission in this regime appears to depend sensitively on the roundness of the bubble immediately before collapse. The same is true for the second phase interval around φ =331° where light emission is observed upon the delayed, strong collapse (Fig. 26).

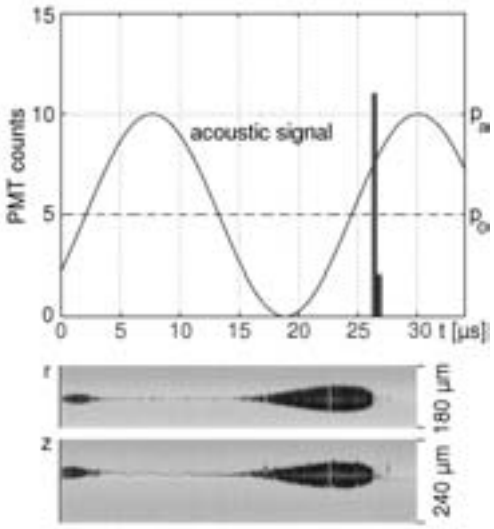


Figure 26. Dynamics and luminescence of bubbles generated at the seeding phase φ_s =331°. The measurements are presented as in Fig. 25. Here, the bubbles are shot in the positive-pressure phase of the sound field and collapse immediately. Upon the second, strong collapse light is emitted.

Nearly perfectly round bubbles can be obtained when laser-breakdown is initiated in the positive sound pressure phase, so that the generated bubble remains initially small and is finally expanded in the subsequent half cycle. The laser energy is largely dissipated, and so are the initial pressure and flow anisotropies induced by the elongated shape of the laser plasma.

This method for bubble inception has been used to study bubble luminescence as a function of acoustic driving pressure. One result is shown in Fig. 27. It presents the luminescence pulse energy (in terms of photomultiplier voltage) vs. acoustic pressure. Measurements were restricted to the first oscillation cycle after bubble generation because, at large driving pressures, the bubbles tend to disintegrate after the collapse and are thus not well defined anymore. The figure shows that with increasing pressure the probability to observe luminescence for a certain laser shot tends toward 100%, and that the average luminescence pulse energy increases as well. It can be concluded that, in fact, by increasing the driving amplitude the bubble collapse becomes more energetic. Above $p_a \approx 4$ bar, which is near the cavitation threshold in the cuvette, the light yield does not increase anymore. The reason for this behaviour remains to be clarified. It may be caused by a shape instability of the bubble, and by the fact that for a fixed frequency of the driving field the bubble can only be expanded to a certain maximum size within one half-cycle of the oscillation.

To observe a larger number of oscillation cycles of laser-generated bubbles in a sound field the acoustic amplitude must be chosen in a way not to excite surface oscillations. This is the case for sound levels around 1.2 bar as used in SBSL experiments. In fact, if the gas content of the liquid is also adjusted correctly, the transition of a laser-generated bubble towards a single stably oscillating, light-emitting bubble can be studied. Figure 28 presents but one example of measurements performed on this phenomenon. Here, the light emission is recorded by means of a sensitive, gatable ICCD camera. The camera was opened for 2.28 ms (corresponding to 100 oscillation

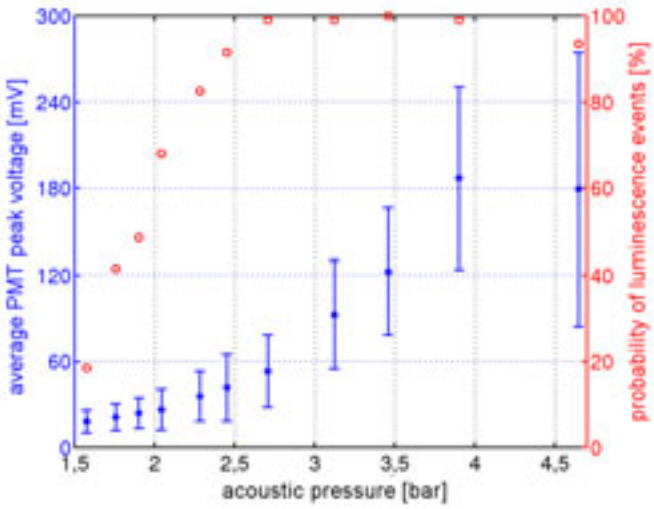


Figure 27. Intensity (∗, blue colour) and relative probability (∘, red colour) of luminescence events at first collapse of a fs-laser-generated bubble in a sound field vs. the acoustic driving pressure. With increasing pressure, the probablity of observing luminescence reaches nearly 100%, while the average energy of the actually observed luminescence events increases up to a sound pressure of about 4 bar, where it appears to level off. All the bubbles were shot in the positive pressure phase of the sound field (measurements by T. Wilken).

cycles) to accumulate photons emitted by the bubble at various time delays after the laser shot. The bubble must be generated at the exact pressure antinode of the sound field to keep it stationary for the time of the measurement (up to several seconds). The figure reveals a steady increase in the emitted number of photons with time on a scale of several seconds. Slow chemical and diffusive processes are presumably taking place during the transition from the initial state to the final SBSL state. The figure also shows that the bubble, at certain times in this process, becomes positionally unstable and performs a translatory or dancing motion. This effect is usually attributed to diffusive instability associated with the shedding of microbubbles. Here, this type of behaviour is observed at a constant driving pressure below the threshold of diffusion instability and could indicate a chemically-induced mass transport.

5 Inside the bubble – Molecular dynamics calculations

During the final stages of collapse the medium within the bubble can become highly compressed and heated, a visible sign of this state being cavitation luminescence. During a time on the order of nanoseconds the liquid inflow is arrested and reversed, and though most models of the radial bubble dynamics consider the bubble medium as an adiabatically and homogeneously compressed ideal or real gas, this is presumably not the true story. To understand cavitation luminescence, and also, for example, bubble chemistry one has to know what is going on inside the bubble.

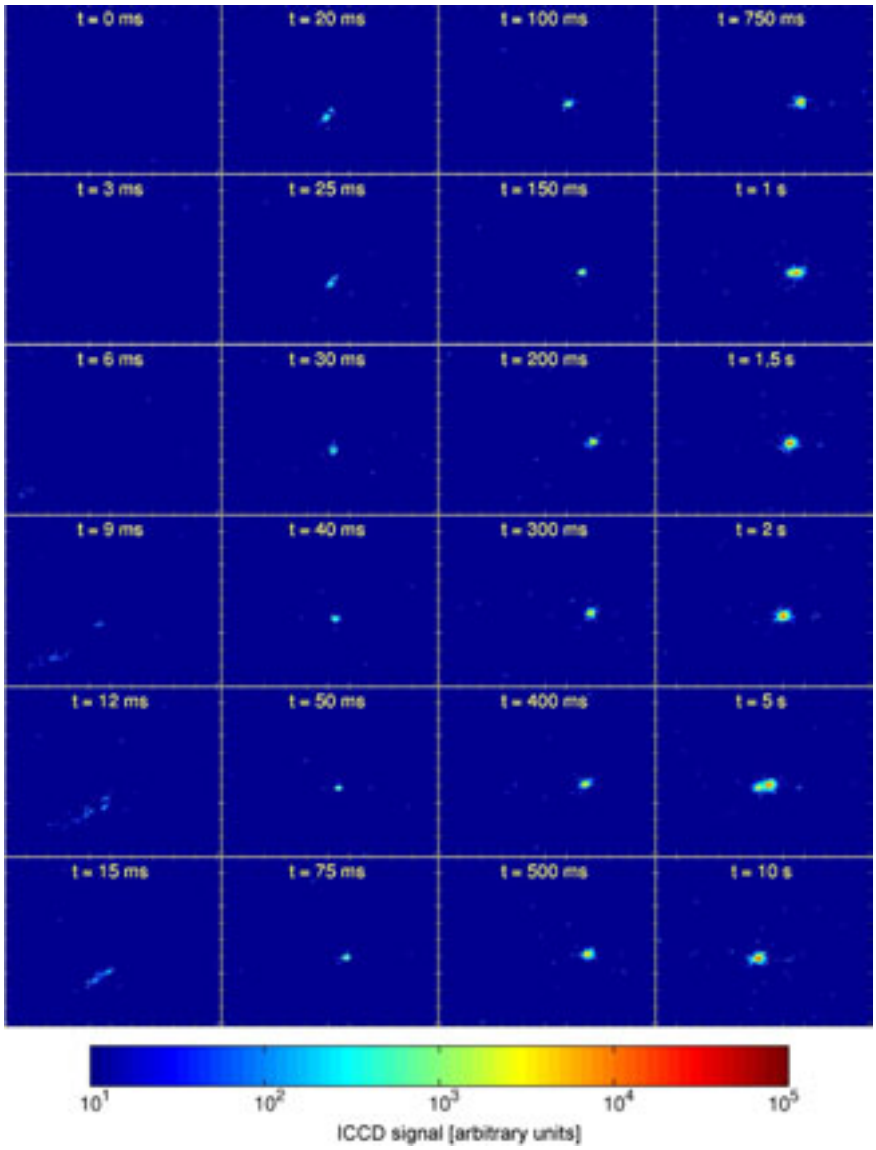


Figure 28. Images of bubble luminescence taken by a gated ICCD camera. Each pixel value represents an average over 5 recordings, the exposure time per recording being 2.28 ms corresponding to 100 oscillation cycles. Bubbles were shot by a femtosecond laser in a sound field ($f_a = 44\,\mathrm{kHz}$) at conditions where stable SBSL is possible (partially degassed water, sound amplitude around 1.3 bar). The times after laser breakdown are given for each frame. The results clearly demonstrate how the laser-generated bubbles slowly turn into a stable SBSL bubble. This process is accompanied by a steady increase of the cavitation luminescence up to full SBSL intensity (measurements by T. Wilken).

At present we are not able to look into the interior of bubbles at collapse to measure pressures and temperatures there, notably at the center. However, numerical investigations can be done to get an impression of the phenomena to expect and to couple them to the observed outside effects. One speculation that has been around for quite a long time [28] is that converging microshocks are launched in a bubble collapsing at supersonic speed, which are held responsible for the extreme heating. After the discovery of SBSL these ideas have been reconsidered and elaborated upon. Initial fluid-dynamical calculations revealed astonishing peak temperatures at the bubble center (greater than 1 million Kelvin [29]). Such results are now considered obsolete, and more sophisticated models [30–32] have yielded an interior dynamics that features inward-travelling compression waves, mixture segregation, vapour trapping, chemical reactions and entropy effects, a part of which may actually prevent the formation of focussing shock waves and limit the peak temperatures.

Notwithstanding theses successes the numerical modelling of the bubble medium remains a formidable problem. A host of physical and chemical mechnisms have to be taken into account, while the bubble medium undergoes extreme changes of its thermodynamic state. The equation of state, chemical reactions and transport phenomena within the medium and across the phase boundary (heat and mass diffusion, phase change) have to be captured. It is clear that detailed models become complicated and difficult to solve, much more difficult to validate. Concerning the question of shock formation, continuous hydrodynamic solvers have to be used that are able to handle the shock discontinuity. This may become problematic for converging shocks whose real thickness near the origin approaches the size of the region to be modelled.

To circumvent this problem, molecular dynamics (MD) simulations of the bubble medium have been proposed [33,34] as an alternative numerical approach. In fact, micrometer-sized bubbles contain relatively few (i. e., on the order of 10^{10}) particles. With suitable coarse-graining, the number of particles (pseudo-molecules) that can be expected to give a faithful representation of reality can be reduced to about one million, a number that is tractable by present-day computers.

In our MD model the gas and vapour molecules are treated as hard spheres that bounce around in the bubble. The hard-sphere system can be advanced in time very efficiently by an event-driven algorithm [35]. The bubble wall is prescribed as a spherical container whose radial dynamics is governed by a Rayleigh–Plesset-type differential equation. It is coupled to the MD model by the pressure exerted on the wall by particle impacts. The physics at the phase boundary is implemented by an appropriate set of wall collision rules [34].

The MD approach naturally includes linear and nonlinear diffusive processes as heat conduction or molecular species diffusion. Chemical reactions can be modelled relatively easily by introducing different particle species and modifying collision rules to reflect possible chemical reactions. Being inherently three-dimensional the general case of aspheric collapse can be simulated without redesign of the algorithm when the bubble wall dynamics is known or can be calculated.

Figure 29(top) depicts the temperature in the interior of a typical sonoluminescing bubble of radius $R_n = 4.5\mu$m driven at 26.5 kHz and 130 kPa as obtained with a MD simulation with 10^7 particles. The bubble contains argon and water vapour. Water dissociation and corresponding chemical reactions have been included in the

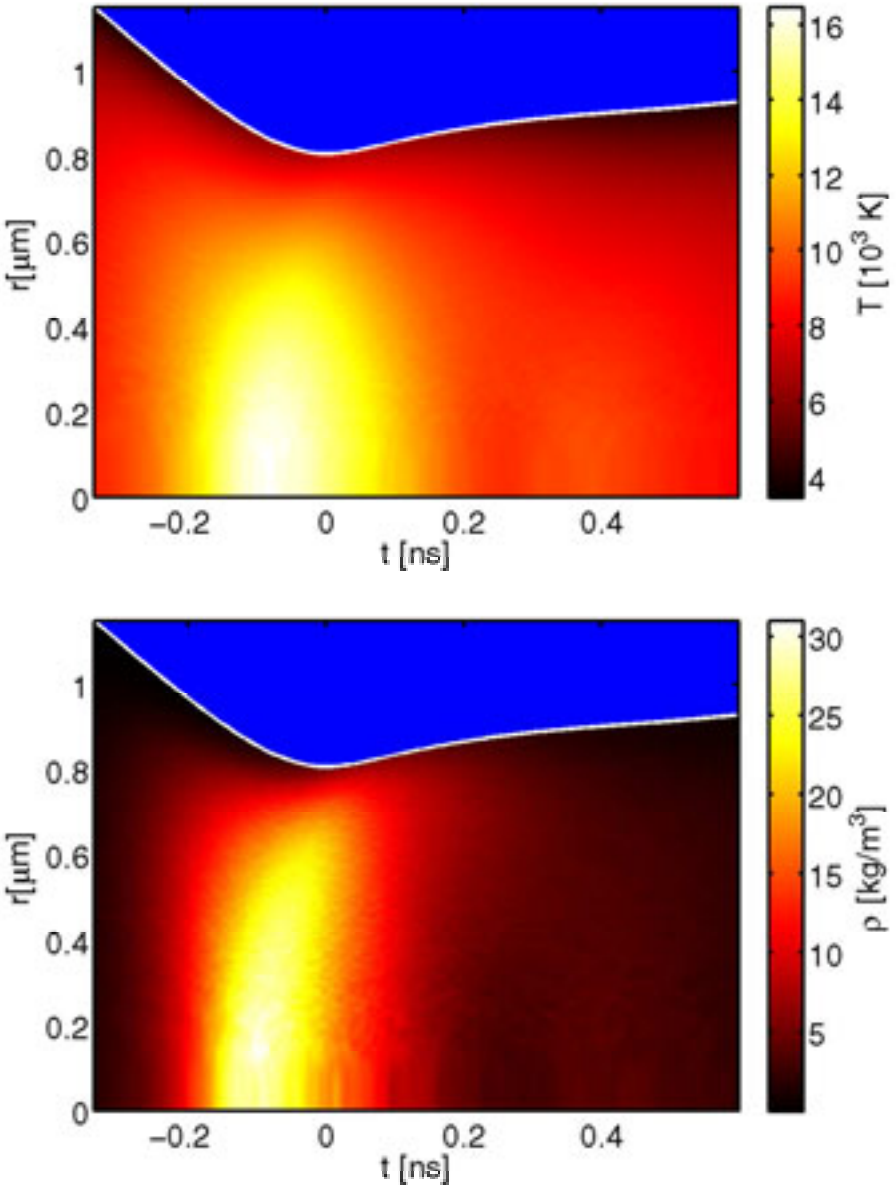


Figure 29. Colour-coded plots of the spatio-temporal evolution of the temperature (top) and the OH density (bottom) in the interior of a collapsing sonoluminescing bubble. Rest radius $R_n = 4.5\mu$m; sound frequency 26.5 kHz; sound pressure amplitude 130 kPa, water temperature 300 K, accomodation coefficient for water vapour $\alpha_v = 0.1$. Ten million particles have been used in the simulation.

calculations. The white line marks the bubble wall. The center temperature reaches about 16000 K upon collapse where the bubble attains a minimum value of about $0.8\,\mu$m. This is in the range of values reported for sonoluminescing bubbles derived from optical spectra of the light emitted. It is seen that the temperature distribution is very inhomogeneous. The density in the bubble turns out to come close to or even exceed that of liquid water.

Figure 29(bottom) presents an example for the capability of the MD method to investigate bubble chemistry. For the same bubble as before the density of OH radicals is plotted in dependence on space and time. The hydroxyl radical is an important chemical species in sonochemistry that can dissolve in the water and mediate further reactions there. The OH distribution closely matches the temperature field. Obviously it is also produced in an outgoing compression wave heading toward the bubble wall. As the temperature drops the OH is quickly consumed again by chemical reactions. Furthermore, in this spherically symmetric collapse the OH concentration remains small in the cold gas-vapour layer near the bubble wall. This means that OH uptake by the liquid should be small as well. It is difficult to accurately predict the amount of OH going into the liquid because the uptake coefficients are not (or not well) known.

The production of OH is closely correlated with the distribution of water vapour in the bubble. It has been shown that in the late collapse phase the water vapour cannot diffuse out of the bubble any more and remains trapped [32,36]. Also, in the presence of non-condensable gas, e. g. noble gases, demixing occurs so that the lighter and hotter molecules accumulate at the center of the bubble. MD simulations confirm this effect and also yield measurable effects in bubble chemistry caused by the redistribution of water vapour.

When fewer particles are used in a molecular dynamics calculation the bubble dynamics can be followed over several oscillation cycles with acceptable computation time. Even with as few as 50000–100000 particles the radial dynamics obtained closely matches results obtained with conventional ODE bubble models (Fig. 30). The advantage is that the MD calculation also provides information on the vapour content and the evolution of chemical species in the bubble over this long time.

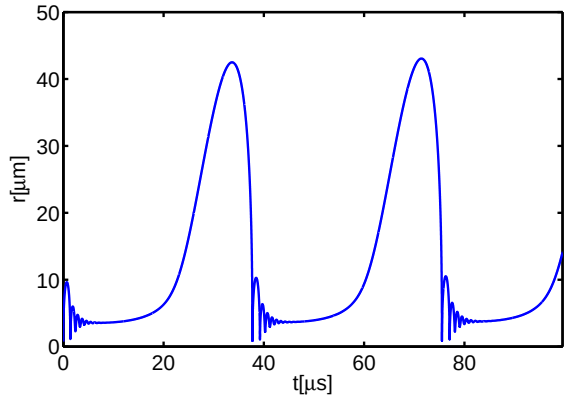


Figure 30. Radius vs. time curve for an SBSL bubble with the same parameters as in Fig. 29. The MD simulation was run with 110000 particles.

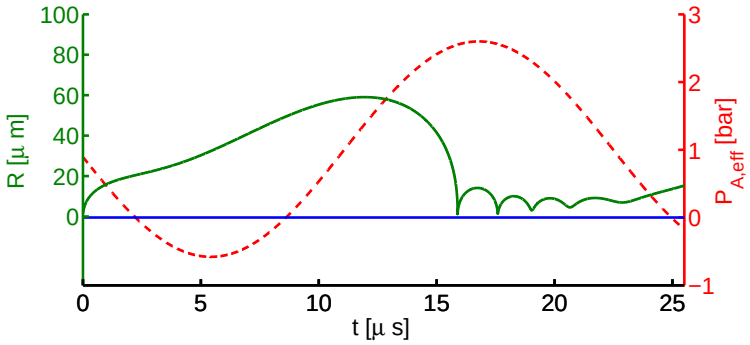


Figure 31. Radius vs. time curve of a laser-generated bubble in an acoustic field of frequency 44 kHz and amplitude 159 kPa. The seeding phase of the bubble is $\varphi_s = 184°$. The MD simulation was run with 64000 particles. The simulation corresponds to the measurement presented in Fig. 25(a).

Of course, the MD method is not restricted to the study of single, sonoluminescing bubbles. The method has also been applied to simulate the interior dynamics of laser-generated bubbles. A particular challenge, and a field for further exploration, is the definition of correct initial conditions to describe an expanding, hot bubble immediately after laser breakdown. In the calculation presented in Fig. 31 for a laser bubble in a sound field, initial conditions were chosen that are believed to be reasonable and consistent with experimental observations of the laser plasma and emerging bubble: the initial radius of 1.5 μm corresponds to the extension of the laser plasma, and the initial temperature was taken as T_0 =3000 K with one third of the water molecules being dissociated. The results on this figure correspond to the experimentally investigated bubbles shown in Fig. 25(a). Even a casual inspection with the naked eye shows that the simulation captures the essential features of the radial dynamics observed in the experiment.

6 Conclusion

The single bubble in a liquid is a conceptually simple two-phase system that nevertheless is difficult to tackle in reality. A multitude of physical and chemical effects occur simultaneously and interact while the bubble undergoes possibly extreme changes in its thermodynamic state, in particular at collapse. Instabilities of position and shape can make it difficult to observe single bubbles experimentally. As we hope to have shown experiments have profited, in this respect, from the method of optic cavitation and the discovery of stable single bubble trapping.

In real-world situations and applications, such as ultrasonic cleaning and chemical processing, bubbles usually appear in clouds and experience strong interaction with each other and with their environment, as clearly demonstrated by a host of self-organization effects [37]. In these settings the bubbles usually have a limited lifetime and undergo changes of size and composition. Nevertheless, they can often be viewed

as the elementary particles in problems of bubble cloud dynamics and structure formation. Single bubble studies as presented here have clarified many basic properties of these objects, and will do so in the future. Still, a lot remains to be learned about single bubbles, in particular, at the nano-scale.

Acknowledgements. The authors thank all current and former members of the Nonlinear Dynamics and Cavitation group at the Dritte Physikalische Institut (DPI) for uncountable input to the work described here, in particular, R. Mettin and U. Parlitz for steady support and scientific exchange, and P. Koch, O. Lindau, C. D. Ohl, and A. Phillipp who contributed to this work. Thanks also go to the wonderful DPI as a whole forming the basis and home where results like those reported here easily could grow.

References

[1] F. D. Gaitan, L. A. Crum, C. C. Church, and R. A. Roy, 'Sonoluminescence and bubble dynamics for a single, stable, cavitation bubble', J. Acoust. Soc. Am. **91**, 3166 (1992).
[2] L. Rayleigh, 'On the pressure developed in a liquid during the collapse of a spherical cavity', Phil. Mag. (Ser. 6) **34**, 94 (1917).
[3] B. E. Noltingk and E. A. Neppiras, 'Cavitation produced by ultrasonics', Proc. Phys. Soc. Lond. **B63**, 674 (1950).
[4] H. Poritsky, 'The collapse or growth of a spherical bubble or cavity in a viscous fluid', in *Proc. First U.S. Nat. Congr. Appl. Mech.*, edited by E. Sternberg (New York, 1952), pp. 813–821.
[5] M. S. Plesset and A. Prosperetti, 'Bubble dynamics and cavitation', Annu. Rev. Fluid Mech. **9**, 145 (1977).
[6] W. Lauterborn, 'Investigation of nonlinear oscillations of gas bubbles in liquids', J. Acoust. Soc. Am. **59**, 283 (1976).
[7] F. R. Gilmore, Tech. Rep. No. 26-4, Office of Naval Research, Hydrodynamics Laboratory, California Institute of Technology, Pasadena, California (1952).
[8] R. Löfstedt, B. Barber, and S. Putterman, 'Toward a hydrodynamic theory of sonoluminescence', Phys. Fluids A **5**, 2911 (1993).
[9] W. Lauterborn, T. Kurz, and I. Akhatov, 'Nonlinear Acoustics in Fluids', in *Springer Handbook of Acoustics*, edited by T. D. Rossing (Springer, New York, 2007), chap. 8, pp. 257–297.
[10] U. Parlitz, V. Englisch, C. Scheffczyk, and W. Lauterborn, 'Bifurcation structure of bubble oscillators', J. Acoust. Soc. Am. **88**, 1061 (1990).
[11] C. Scheffczyk, U. Parlitz, T. Kurz, W. Knop, and W. Lauterborn, 'Comparison of bifurcation structures of driven dissipaptive nonlinear oscillators', Phys. Rev. A **43**, 6495 (1991).
[12] W. Lauterborn and U. Parlitz, 'Methods of chaos physics and their applications to acoustics', J. Acoust. Soc. Am **84**, 1975 (1988).
[13] J. B. Keller and M. Miksis, 'Bubble oscillations of large amplitude', J. Acoust. Soc. Am. **68**, 628 (1980).
[14] Y. Tian, J. A. Ketterling, and R. E. Apfel, 'Direct observation of microbubble oscillations', J. Acoust. Soc. Am. **100**, 3976 (1996).
[15] J. Holzfuss, M. Rüggeberg, and A. Billo, 'Shock wave emissions of a sonoluminescing bubble', Phys. Rev. Lett. **81**, 5434 (1998).
[16] B. Gompf, R. Günther, G. Nick, R. Pecha, and W. Eisenmenger, 'Resolving sonoluminescence pulse width with time-correlated single photon counting', Phys. Rev. Lett. **79**, 1405 (1997).

[17] M. P. Brenner, S. Hilgenfeldt, and D. Lohse, 'Single-bubble sonoluminescence', Rev. Mod. Phys. **74**, 425 (2002).
[18] D. Lohse, M. P. Brenner, T. F. Dupont, S. Hilgenfeldt, and B. Johnston, 'Sonoluminescing Air Bubbles Rectify Argon', Phys. Rev. Lett. **78**, 1359 (1997).
[19] W. Lauterborn, 'Optische Kavitation', Phys. Bl. **32**, 553 (1976).
[20] W. Lauterborn and H. Bolle, 'Experimental investigations of cavitation-bubble collapse in the neighbourhood of a solid boundary', J. Fluid Mech. **72**, 391 (1975).
[21] A. Vogel, W. Lauterborn, and R. Timm, 'Optical and acoustic investigations of the dynamics of laser-produced cavitation bubbles near a solid boundary', J. Fluid Mech. **206**, 299 (1989).
[22] A. Philipp and W. Lauterborn, 'Cavitation erosion by single laser-produced bubbles', J. Fluid Mech. **361**, 75 (1998).
[23] O. Lindau and W. Lauterborn, 'Cinematographic observation of the collapse and rebound of a laser-produced cavitation bubble near a wall', J. Fluid Mech. **479**, 327 (2003).
[24] E. Zwaan, S. Le Gac, K. Tsuji, and C. D. Ohl, 'Controlled Cavitation in Microfluidic Systems', Phys. Rev. Lett. **98**, 254501 (2007).
[25] K. Tsiglifis and N. A. Pelekasis, 'Nonlinear oscillations and collapse of elongated bubles subject to weak viscous effects', Phys. Fluids **17**, 102101 (2005).
[26] C. D. Ohl, O. Lindau, and W. Lauterborn, 'Luminescence from Spherically and Aspherically Collapsing Laser Induced Bubbles', Phys. Rev. Lett. **80**, 393 (1998).
[27] T. Kurz, D. Kröninger, R. Geisler, and W. Lauterborn, 'Optic cavitation in an ultrasonic field', Phys. Rev. E **74**, 066307 (2006).
[28] P. Jarman, 'Sonoluminescence: A Discussion', J. Acoust. Soc. Am. **32**, 1459 (1960).
[29] C. Wu and P. H. Roberts, 'Shock-Wave Propagation in a Sonoluminescing Gas Bubble', Phys. Rev. Lett. **70**, 3424 (1993).
[30] W. Moss, A. Young, J. Harte, J. Levatin, B. Rozsnyai, G. Zimmerman, and I. Zimmerman, 'Computed optical emissions from a sonoluminescing bubble', Phys. Rev. E **59**, 2986 (1999).
[31] B. D. Storey and A. J. Szeri, 'Mixture segregation within sonoluminescing bubbles', J. Fluid Mech. **396**, 203 (1999).
[32] B. D. Storey and A. J. Szeri, 'Water vapour, sonoluminescence and sonochemistry', Proc. Roy. Soc. Lond. A **456**, 1685 (2000).
[33] B. Metten and W. Lauterborn, 'Molecular Dynamics Approach to Single-Bubble Sonoluminescence', in *Nonlinear Acoustics at the Turn of the Millenium: Proc. 15th Int. Symp. on Nonlinear Acoustics*, edited by W. Lauterborn and T. Kurz (American Institute of Physics, Melville, New York, 2000), p. 429.
[34] B. Metten, *Molekulardynamik-Simulationen zur Sonolumineszenz* (Der Andere Verlag, Osnabrück, 2001).
[35] D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge University Press, Cambridge, 2004), 2nd ed.
[36] R. Toegel, B. Gompf, R. Pecha, and D. Lohse, 'Does Water Vapor Prevent Upscaling Sonoluminescence?', Phys. Rev. Lett. **85**, 3165 (2000).
[37] R. Mettin, 'From a single bubble to bubble structures in acoustic cavitation', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).

Oscillations, Waves and Interactions, pp. 171–198
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-07-3

From a single bubble to bubble structures in acoustic cavitation

Robert Mettin

Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. The article reviews main ingredients for a bottom-up approach to understand and model acoustic cavitation structures. Based upon the dynamics of a single, spherical bubble in a sound field, the fate of a bubble in a given acoustic set-up and collective behaviour in multibubble systems are derived. The discussion includes nucleation and oscillation of bubbles, their shape stability, and rectified gas diffusion. Furthermore, acoustic forces on pulsating bubbles and bubble translation are addressed. The different aspects are combined in bubble life cycle diagrams, and a particle model for the description of many interacting bubbles is proposed. The application of this method for numerical simulation of experimentally observed acoustic cavitation structures is demonstrated by examples.

1 Introduction

The expression *acoustic cavitation* describes the physical phenomenon of rupture of a liquid under tension caused by a sound wave [1–6]. Apart from the creation of voids in the liquid itself, which is termed *inception*, acoustic cavitation comprises as well the further destiny of the voids. This leads directly to the main subjects *bubble dynamics*, *bubble collapse*, and *bubble structure formation*. An important issue is also the transformation of the non-cavitating liquid into a *two-phase fluid* of gas and liquid after cavitation inception. Among other effects, this can cause a *retroaction* of cavitation on the generating sound field.

All the above items, taken alone, are already challenging research topics and not understood in every detail yet. Taken all together, we can rightly characterize acoustically cavitating liquids as *complex systems*. Complicating aspects for a general theoretical treatment of acoustic cavitation and the accompanying phenomena are especially *nonlinearity*, time and space *scales* that span many orders of magnitude, and *granularity* (i. e., properties somehow intermediate between continuous and discrete). Accordingly, investigations in this area often attack partial problems and idealized cases to reach a better and better understanding and description of the observations, though still being far from a general and complete picture.

In recent time, the demand for an improved knowledge of acoustic cavitation has grown constantly, because the field of applications is permanently extending. Be-

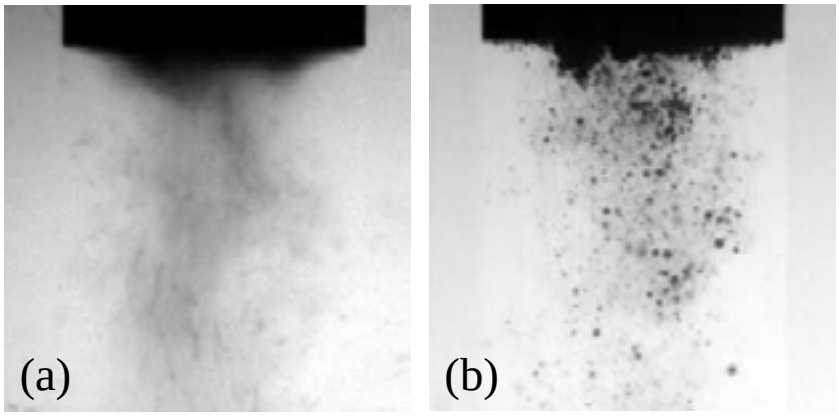


Figure 1. Bubbles in tap water below a sonotrode tip (5 mm diameter, 20 kHz) with background illumination: exposure time in (a) 1 ms, in (b) 1 μs.

sides rather established usage in ultrasonic cleaning and ultrasonically assisted chemistry, relatively new and even exotic applications gain attention. These range from nanoparticle production and sewage sludge activation via biomedical applications like contrast imaging and cancer treatment up to the prospect of table-top nuclear fusion reactions. In parallel with the increased interest, research methods and technical equipment have improved likewise in recent years. This is true in particular for high-speed imaging systems and computer power. Consequently, research activity is very active and growing, and new insights are obtained rapidly.

The aim of this article is to focus on the destiny of individual bubbles in acoustic cavitation. The acoustic field causes a permanent driving of the once created cavity, and this leads to peculiar effects that prolongate and complicate the life of a bubble. One important microscopic aspect is the bubble collapse [7]. On a meso- and macroscopic scale, it is important that the bubbles often translate quite fast due to acoustic forces.[1] Additionally, they can shrink or grow by gas diffusion processes, merge with other bubbles, or split into smaller ones, and these processes are also strongly influenced by the acoustic field.

In typical set-ups of cavitation applications, an acoustic field is created by an ultrasonic transducer. Here, one can roughly distinguish between localized sources, like the tip of an ultrasonic horn, and plane emitters of larger size (compared to the acoustic wavelength). In any case, cavitation manifests itself to the unaided observer as "clouds" or "bubbly zones" of various shapes and extensions. Of course, the cavitating region consists of many individual, often well separated and surprisingly spherical bubbles. This is even true at rather high bubble densities, as is demon-

[1]This might be seen in contrast to *hydrodynamic cavitation*, where the liquid ruptures under a hydrodynamic flow condition which induces negative pressures. Then, cavities are typically attached to a body surface, or are convected by the flow to normal pressure zones, and disappear. Therefore, the life of a hydrodynamic cavitation bubble is usually quite different, and its location is determined by the body and/or the liquid flow.

strated in Fig. 1 by two photographs of long and short exposure time, respectively. This clearly motivates the description of the cavitating zone via many single bubbles and their respective behavior. For the consideration of spatial bubble distributions, we can treat them as individual *point-like particles*: each bubble has a well defined position and velocity, and it is reacting to forces acting upon its centre of mass. Following this approximation, the dynamics of the two-phase liquid transfers to a many-body problem, in a way similar to stars moving in a galaxy, or electrons and ions moving in a plasma.[2] Indeed, notable parallels exist as compared to gravitation or electromagnetism, but the acoustic origin of the forces and the presence of a liquid bear also specific differences, which renders the many-body problem a rather peculiar one.

The following Sections 2 and 3 will briefly discuss the origin of bubbles and their oscillation dynamics under the presence of a driving sound field. The treatment is limited to spherical bubbles, and the loss of spherical stability as well as rectified gas diffusion are addressed. Section 4 deals with acoustic forces on the bubbles and aspects of their translational motion like added mass and viscous drag. A brief discussion of bubble life cycles follows in Sect. 5. The extension to a multibubble particle model is proposed in Sect. 6, and some examples of spatial bubble structure formations together with their numerical simulation are presented. A conlusion and outlook is given in Sect. 7.

2 Cavitation inception and bubble nucleation

If a continuous (cw) acoustic field of gradually increased intensity is applied to a liquid, sooner or later the cavitation inception threshold will be reached, and bubbles are created. Because the sound is permanently irradiated into the now cavitating liquid, generated bubbles will have a subsequent (possibly complicated) destiny, and further new bubbles can appear all the time. Reserving the term "inception" for the global transition from quiet to cavitating liquid, here the notion of "bubble creation" will be preferred for the local origin of a bubble. Indeed, one can observe several differerent ways of bubble creation in cw fields. Some bubbles appear suddenly "out of nothing" in the bulk liquid, others seem to occur from an invisible continuous source in the bulk, or from continuous sources at walls. Sometimes the source is a larger bubble that emits smaller ones. The nature of bubble creation or source can be very important for the spatial cavitation patterns that appear, and some are illustrated in the following.

Figure 2 shows the process of a spontaneous nucleation near a high acoustic pressure zone in bulk water. A small object suddenly appears isolated and without any precursor in the liquid (first frame) and develops into a cluster. It travels at relatively high speed towards a lower acoustic pressure region. Crum and Nordling

[2] Another point of view is taken by *continuous models* of cavitation, where the gas void fraction or the density of bubbles is understood as a quantity continuously varying in space and time. Such models do not resolve individual bubbles, and they are usually formulated by partial differential equations. A particle approach results in a set of ordinary differential equations.

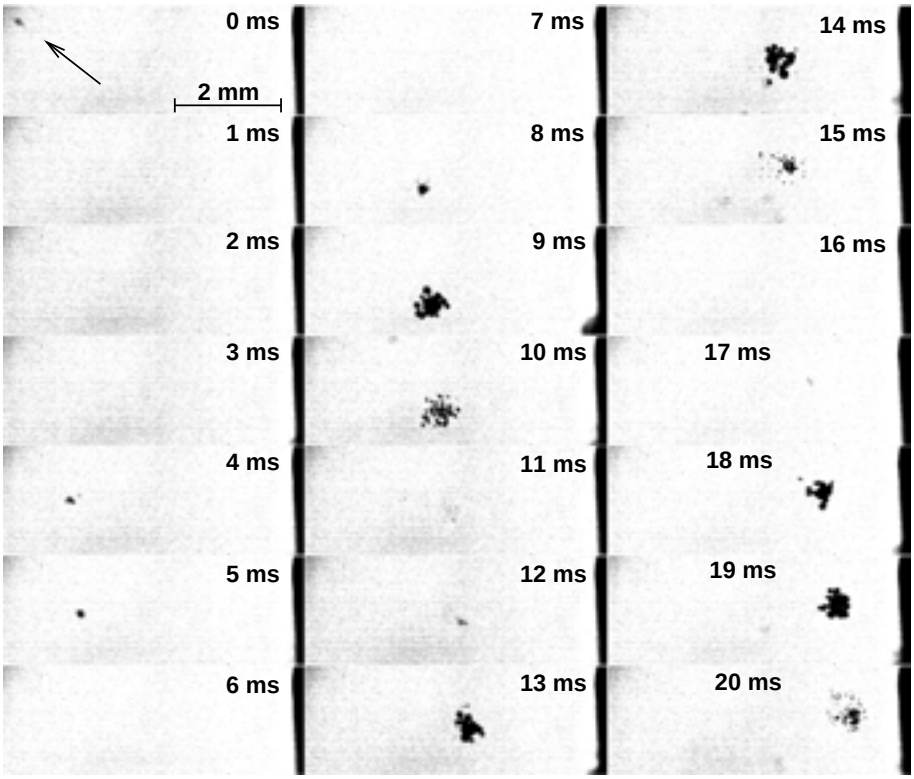


Figure 2. Free nucleation into a cluster in a 20 kHz sound field (high-speed recording with 1000 frames per second (fps), exposure time 1.5 μs). An object that appears as a germ bubble (arrow first frame) develops into a conglomerate of bubbles that travels to the right. The visibility of the cluster changes, because the relative phase of the sub-period exposure time is drifting from frame to frame, and the cluster bubbles oscillate with strong volume changes. On some frames they are close to collapse and too small to be resolved.

observed similar events [8] and termed them "comets". Possibly they are triggered by an advected microbubble, or ionizing particle radiation.

Another typical picture of nucleation is shown in Fig. 3. A quasi-continuous stream of small bubbles appears out of the bulk, forming a trail of bubbles, or a "streamer". Bubbles travel along directional paths from the invisible source in the liquid towards junctions with other streamers, which causes a dendritic form. The invisible sources of bubbles remain active over many acoustic cycles. Probably, invisible sub-micron bubbles, stabilized or from degassing processes in the liquid, migrate to the apparent bubble source and merge to visible size, feeding the stream of bubbles.

Further sources of small bubbles are larger ones: driven by the acoustic field, they can undergo surface instabilities. By pinch-off or even by apparent "explosions" [9] one or many daughter bubbles can be produced. The generating bubble may be freely floating in the liquid, or be attached to an immersed object or a wall. If fixed,

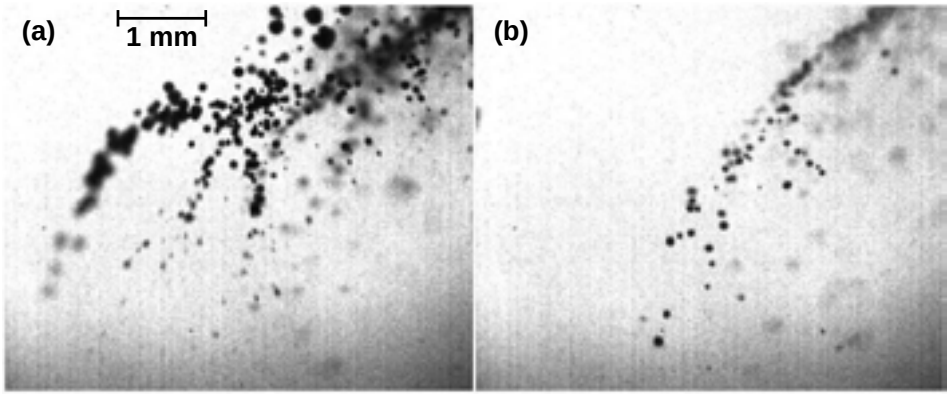


Figure 3. Nucleation from the bulk liquid into streamers in a 20 kHz sound field (two frames from a high-speed recording, exposure time 1.5 μs). Several streamer sources on frame (a) change into a single one on frame (b) within the interframe time of 30 ms.

it can form a spatially stable source of a continuous bubble stream. To control such attached bubble sources can be one way of a cavitation structure control [10].

Examples of microbubble generation are shown later in Figs. 7 and 9. The process of repetitive ejection seems to be a common cavitation bubble source if objects or walls are contributing, and for a complete description, gas diffusion into such source bubbles has to be taken into account. The attached bubbles might even be very small, as recent investigations on "nano-bubbles" suggest [11].

3 Bubble dynamics

Once a bubble is created in the liquid, it will show variations of size and possibly shape as a reaction to the sound field. On a longer time scale, translational and diffusional processes can take place. From the point of view of a single, individual bubble, these issues are captured by the term *bubble dynamics*. A very useful model for investigation of bubble dynamics is a *spherical gas and vapour bubble*. Here it is usually assumed that the gas inside the bubble is non-condensable (and thus its mass is conserved, at least on a fast time scale), and the vapour is instantaneously evaporating or condensing at the bubble walls, leading always to the equilibrium vapour pressure inside.[3] The presence of a non-condensable gas (in many cases air) can be questioned, but keeping in mind the nucleation mechanism from stabilized micro (or nano) gas bubbles, this assumption usually makes sense.[4] Indeed, an

[3]Refinements of this model include gas diffusion and vapour evaporation and condensation dynamics, but they lead in many parameter regimes in water to rather small corrections. They are important for a good description of the collapse, which is not central here, and thus they are neglected in the following.

[4]Of course, also *pure vapour cavitation* might be encountered in some situations, like in liquid helium [12]. However, here we are dealing mainly with water-air systems under normal conditions.

important quantity characterizing every bubble is the mass of non-condensable gas inside, defining its *equilibrium size*. On a longer time scale, this size can change due to diffusion (see below, Sect. 3.3).

The spherical symmetry renders the problem one-dimensional and facilitates calculations significantly. Indeed, this assumption is surprisingly good in many situations encountered in acoustic cavitation, and mainly violated for larger bubbles and for very close boundaries imposed by other bubbles or objects. Instability of sphericity is also discussed below (Sect. 3.2).

Spherical gas bubble models mainly vary in the Equation Of State (EOS) for the liquid and the gas. The simplest assumption is an incompressible liquid and an ideal gas, resulting in an equation frequently referred to as the Rayleigh-Plesset equation [4,5]. It can be improved by incorporation of liquid compressibility and an advanced EOS for the gas like a Van-der-Waals model. For the problems considered here, a sufficient compromise between computational effort and accuracy is a model with slight (first order) compressibility of the liquid, namely the Gilmore model [13] or the Keller-Miksis model [14]. With an ideal gas law the Keller-Miksis model reads

$$\left(1-\frac{\dot R}{c}\right)R\ddot R+\frac{3}{2}\dot R^2\left(1-\frac{\dot R}{3c}\right)=\left(1+\frac{\dot R}{c}\right)\frac{p_l}{\rho}+\frac{R}{\rho c}\frac{dp_l}{dt}\,,$$

$$p_l=\left(p_0+\frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\kappa}-p_0-\frac{2\sigma}{R}-\frac{4\mu}{R}\dot R-p_a(t)\,.$$

Here, R is the bubble radius, c and ρ the sound speed and the density of the liquid, μ and σ the viscosity and the surface tension, and κ the polytropic exponent. The acoustic pressure is denoted $p_a(t)$, while the static pressure is p_0. R_0 denotes the bubble equilibrium radius, and vapour pressure is neglected.

In the following we focus on the *volume oscillations*, *shape instabilities*, and *rectified diffusion*. Afterwards, *acoustic forces* are addressed which are induced by a cw acoustic field on a gas bubble. The acoustic forces finally lead to *spatial translation* and *structure formation* of the bubbles. For all calculations, the spherical bubble model is used and thus it is the core of all results presented.

3.1 Bubble oscillations

A spherical gas-filled bubble under static conditions posseses an equilibrium radius R_0 (or an equilibrium volume $V_0=4\pi R_0^3/3$) if the static pressure p_0 is larger than the Blake threshold pressure [4]

$$p_B=p_0+\frac{8\sigma}{9}\left(\frac{3\sigma}{2[p_0+(2\sigma/R_0)]R_0^3}\right)^{1/2}\,.$$

For $p_0<p_B$ the bubble would expand infinitely.[5] A momentary excursion of the bubble radius from the equilibrium results in damped oscillations around R_0, the

[5] However, in practice it is normally not possible to sustain a static negative pressure for a longer time under the condition of bubble expansion, i. e., cavitation. Thus bubble expansion will be stopped sooner or later in real systems.

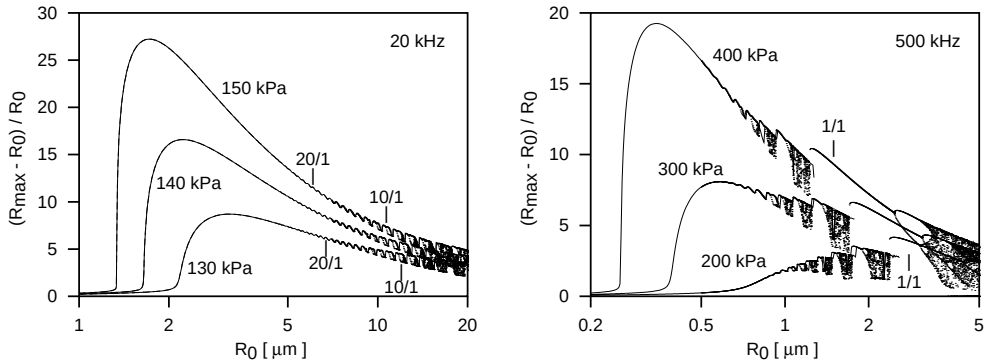


Figure 4. Bubble response at driving frequencies of 20 kHz (left) and 500 kHz (right). The normalized maximum bubble radius during one driving period is plotted vs the equilibrium radius for different driving pressures. If higher-periodic or chaotic solutions occur, more than one point is visible. Certain resonances are indicated by numbers (from Ref. [17]).

restoring force being the static liquid pressure (for $R > R_0$) or the gas pressure (for $R < R_0$). If the pressure oscillates around the static value, which is the case in a sound wave, the bubble radius performs forced oscillations: the gas bubble behaves as a driven oscillator system [5]. Small excitation allows to consider the linearized model, which shows a resonance frequency of [15]

$$\omega_{\mathrm{res}}^2 = (2\pi f_{\mathrm{res}})^2 = \frac{1}{\rho R_0^2}\left(3\kappa p_0 + \frac{2\sigma}{R_0}(3\kappa - 1)\right) .$$

In the case of water under normal conditions, we find the approximate relation $f_{\mathrm{res}} R_0 \approx 3\,[\mathrm{m/s}]$. For monofrequent driving at the frequency f we can accordingly find a resonant equilibrium radius $R_{\mathrm{res}} \approx 3/f\,[\mathrm{m}]$. The linearized pulsation of the gas bubble is important for analytical considerations of many aspects of bubble dynamics and as a reference motion. It is discussed in textbooks [4–6]. However, one has to keep in mind that it is only valid in the limit of small bubble volume changes, and stronger excursions exhibit the nonlinear behaviour of the equation. In particular, nonlinear resonances, hysteresis, and chaotic oscillations [16] occur for suitable parameters. This has been shown for parameters relevant for acoustic cavitation at lower ultrasonic frequencies in Ref. [17]. From there, Fig. 4 has been taken for illustration of the rich behaviour.

A specific and important feature of the bubble response is the "dynamic" Blake threshold: for sufficiently small bubbles, surface tension can not be overcome by the negative acoustic pressure and the bubbles oscillate only weakly. From a certain equilibrium radius on (or similar, from a certain stronger tension on), the bubble grows to a multitude of its equilibrium size and shows subsequent strong collapse and rebound events. The transition or border between small, "quiet" and "inactive" bubbles and the larger, strongly expanding "loud" and "active" bubbles is relatively sharp and equivalent to the standard static Blake threshold in the limit of an acous-

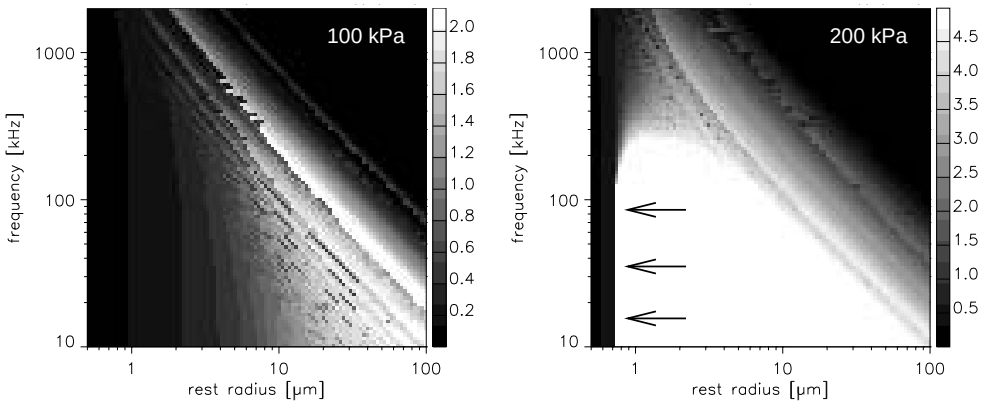


Figure 5. Dynamic Blake threshold in the parameter plane of equilibrium radius R_0 and frequency f. The normalized maximum radius $(R_{\max} - R_0)/R_0$ is gray coded for the two driving pressures 100 kPa (left) and 200 kPa (right). At 200 kPa, the sharp border of the white area constitutes the threshold (arrows).

tic frequency of zero.[6] This phenomenon has been highlighted by Noltingk and Neppiras [18,19]. In more detail it has been calculated and discussed later, for instance in Refs. [17,20], and its relevance for trapped single bubbles and single bubble sonoluminescence (SBSL) has been realized in Refs. [21,22]. Figure 5 from Ref. [20] illustrates the sudden jump in parameter space of equilibrium radius and acoustic frequency. The distinction between weakly and strongly oscillating bubbles, i. e., below and beyond the dynamic Blake threshold, is very important. This is true obviously for applications based on heavy bubble collapse, but also for structure formation of acoustic cavitation bubbles, because acoustic forces, gas diffusion, and stability can be changing dramatically when crossing this border. These issues are addressed in the following sections.

3.2 Spherical shape stability

The spherical shape of a cavitation bubble is idealized. Due to surface tension, a static bubble will always tend to form a sphere, which is also stable with respect to small perturbations. This does not hold necessarily for a pulsating bubble. To estimate the stability of a driven oscillating spherical bubble, usually the bubble shape is expanded into spherical harmonics (modes), and the equations of motion of the corresponding expansion coefficients are considered. For small excursions from the spherical form, the equations can be linearized, and it turns out that then only unidirectional coupling of the fundamental (radial) motion to the higher order surface modes persists (no mutual coupling or back-coupling to the radial oscillation, see Refs. [4,5,23–25]). Formally, in linear approximation all mode coefficients are described by parametri-

[6] The notion "(in)active" relates to all cavitation effects due to a strong bubble collapse. Nevertheless, bubbles below the dynamic Blake threshold oscillate, feel acoustic forces and might be "active" from another point of view.

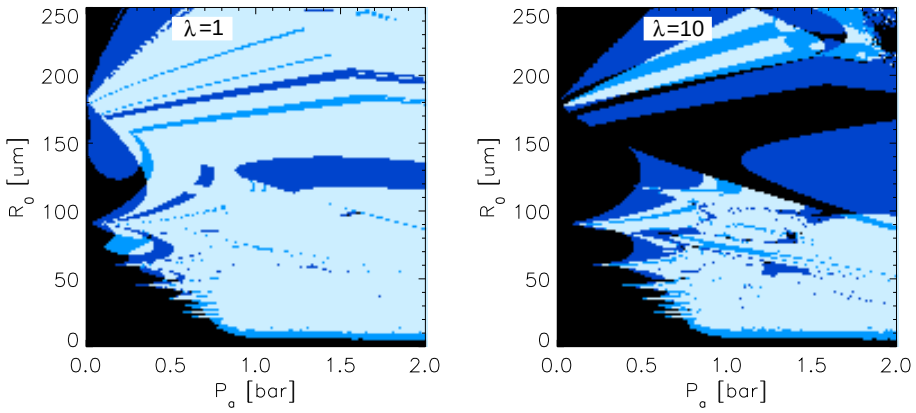


Figure 6. Stability of a spherical bubble with respect to parametric surface mode instability in the parameter plane of equilibrium radius R_0 and driving pressure amplitude p_a; driving frequency 15 kHz. Colors indicate the stability with respect to the modes $n = 2$, 3, and 4. Black: all modes stable, darkest blue: one mode unstable, blue: two modes unstable, and light blue: all three modes unstable. Left: maximum allowed stable eigenvalue $\lambda = 1$, right: $\lambda = 10$. At higher driving, all but very small bubbles (below the dynamic Blake threshold) become unstable.

cally excited linear oscillator systems, which are for periodic radial oscillation of the type of Hill's differential equation. If a_m denotes the amplitude of the m-th surface mode, its equation of motion reads

$$\ddot{a}_m + \left[3\frac{\dot{R}}{R} + 2(m+2)(2m+1)\frac{\mu}{\rho R^2}\right]\dot{a}_m$$
$$+(m-1)\left[(m+1)(m+2)\frac{\sigma}{\rho R^3} - \frac{\ddot{R}}{R} + 2(m+2)\frac{\mu\dot{R}}{\rho R^3}\right]a_m \quad = \quad 0\,.$$

Here, the solution of the spherical bubble model, $R(t)$, enters as the parametric driving. The bubble is considered to be spherically unstable, if at least one surface mode amplitude reaches the actual bubble radius value (*Rayleigh-Taylor instability*[7]) or grows indefinitely in time (*parametric instability*). For simplicity, only the latter case is considered in the following. It is easily calculated numerically by integrating the spherical bubble model together with the linear mode amplitude equations. Because of linearity of the latter, it is sufficient for a periodic radial motion to integrate over one period and calculate the Floquet multipliers (eigenvalues of the "monodromy matrix", see Ref. [26]). If the absolute value of at least one mode is larger than one, formally that mode would grow and finally destroy the bubble. In reality, the linearization breaks down from some point on, and mode coupling would lead to energy redistribution, also to the fundamental mode. Without considering this issue here in

[7]For this calculation, one has to assume an initial size of a disturbance of the spherical form, and the outcome slightly depends on this choice.

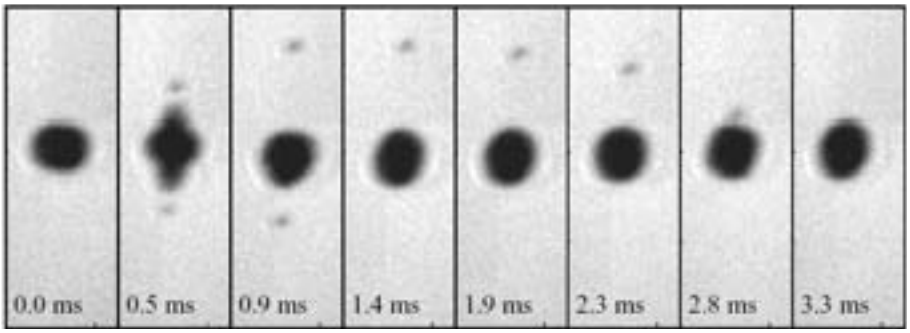


Figure 7. Trapped single bubble shedding microbubbles and recollecting them (high-speed recording at 2250 fps in a 25 kHz standing wave; frame width 1.2 mm; from Ref. [29]).

detail, it is possible to replace the absolute value "1" for an unstable eigenvalue by a higher number, e. g. 10 or 100, to model the larger damping of the mode. Indeed experimentally observed surface instability of a single bubble could be better modelled by this approach of an artificially higher stable eigenvalue [27,28]. An illustration is given in Fig. 6, where the stability regions of several modes for a relevant parameter range are plotted colour coded. Typically, the most unstable surface mode is $n = 2$, and therefore it is often sufficient to calculate its behaviour.

If a bubble is spherically unstable, several things can happen: (1) It might shed smaller bubbles to lose gas, reduce its equilibrium radius accordingly, and thus become more stable (i. e., jump more to the bottom in Fig. 6). Afterwards, it might (1a) regain spherical stability if the fragment disappears, or (1b) recollect the shedded microbubble(s) in a continuing process and thus form a "jittering" or "dancing" bubble, see Fig. 7 for an example. (2) Alternatively the bubble might split into many fragments of similar size. The fragments can subsequently reunite again to form one or more larger bubbles. If the destruction and merging process goes on, a small cluster of fragment bubbles is observed, see Fig. 8. (3) The bubble might get destroyed and vanish. This process is possibly an "atomization" into many very small fragments that rapidly dissolve into the liquid.[8] As the examples have shown, the spherical instability does not necessarily mean a destruction of a bubble, but may at times give rise to a multitude of possibly interacting bubbles and thus to a *bubble structure*. As an extreme example, Fig. 9 shows a quite large bubble attached to a transducer surface, a sonotrode tip. This bubble is long-term existent, but shows strong deviations from (semi)spherical shape. The high-speed sequence reveals that it is also strongly collapsing. It continuously sheds smaller and larger daughter bubbles, but keeps its size by remerging bubbles and possibly by diffusion processes. From a spherical stability point of view, such bubbles should be very short-living, but in some cases as the observed, they live quite long and serve as permanent bubble source (cf. Sect. 2). Of course, the presence of a boundary is of importance here.

[8] This happens typically in bubble trap experiments if the driving pressure is tuned too high: the trapped bubble seems to disappear within a very short time [30].

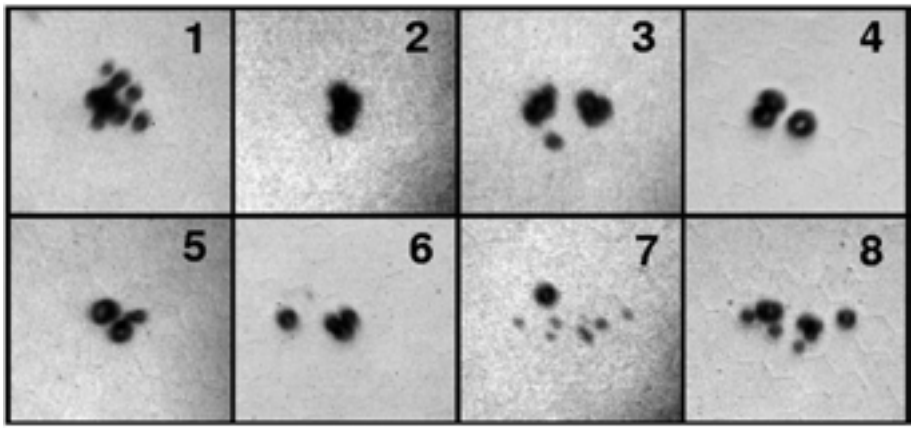


Figure 8. Continuously merging and splitting small cluster of bubbles close to a pressure antinode (high-speed recording with one frame per driving period, i. e., 25000 fps; exposure time 1 μs; frame size approximately 1 mm); from Ref. [10].

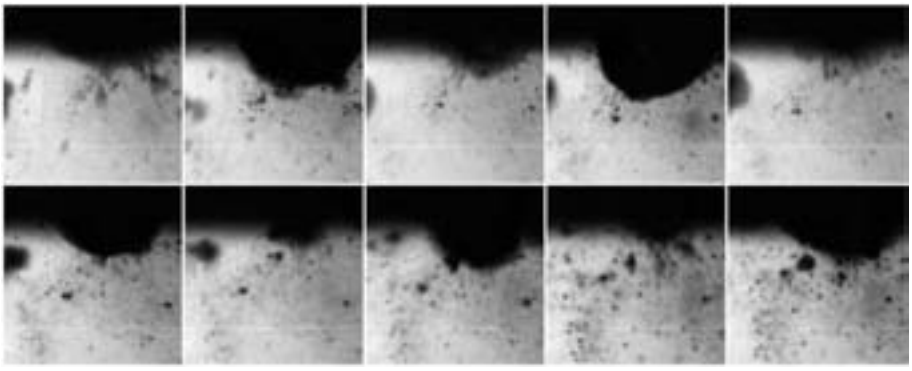


Figure 9. Large bubble at a 20.61 kHz sonotrode tip. The subsequent frames (top left to bottom right) from a high-speed recording at 20000 fps (exposure time 20 μs, frame size approx. 5 mm) show one picture per acoustic cycle. The bubble is roughly half-spherical and strongly oscillating in a subharmonic (period-doubled) way. It is continuously shedding microbubbles and exists much longer than the sequence shows.

Frequently, however, one encounters the cases (1a) and (3) from above, and the bubble sizes are limited by the spherical shape stability threshold. This will be used later for an upper estimate of the sizes of longer living bubbles.

3.3 Rectified diffusion

The mass and the species of non-condensable gas inside a bubble can change due to diffusion through the bubble wall, i. e., from or into the liquid. Typically gases have a finite solubility in liquids, and in dilute solutions at equilibrium, the gas

concentration of species i in the liquid c_i is proportional to the gas partial pressure p_i above the liquid: $c_i = p_i/H_i$ with Henry's constant H_i.[9]

If we consider a static gas bubble in a liquid that is saturated with the gas, the partial pressure inside the bubble will be higher than the static pressure outside because of surface tension. Therefore the gas will go into solution and the bubble will dissolve sooner or later. This process is even accelerated as the bubble shrinks, since the surface tension grows with the inverse bubble size. For small bubbles in the micrometer range, dissolution happens quite rapidly on the scale of milliseconds to seconds [31]. The same holds for undersaturated liquids. Indeed, only for oversaturated liquids a static diffusional equilibrium can be reached, which however turns out to be unstable: bubbles either dissolve, or grow unlimited (and then rise to the liquid surface due to gravity). This behaviour demands for a stabilization mechanism of microbubbles in the context of cavitation nuclei [4,5].

If the bubble is not static, but oscillating in a sound pressure field, the bubble size and the gas pressure change on the acoustic time scale which is typically fast compared to the diffusional time scale. It turns out that the gas diffusion *into* the bubble is favoured in such a case by three mechanisms [5]: (i) the surface effect, i. e., a larger gas-liquid interface at low gas pressure in the expanded bubble cycle; (ii) the shell effect, leading to a larger concentration gradient at the bubble wall during the expanded bubble phase; and (iii) the nonlinear oscillation effect, causing a longer time per period spent at large bubble sizes, i. e., at low inside gas pressures. Mechanisms (i) and (ii) can be understood if one recalls that the diffusion is governed by Fick's law and thus the mass flow rate is proportional to the interface area and the concentration gradient [5]. Mechanism (iii) follows from the unsymmetric potential of the bubble oscillation, representing a hard spring oscillator for compression and a soft spring for expansion [16].

The above effects can overbalance the dissolution tendency given by surface tension, and oscillating bubbles can show a net growth for saturated and even undersaturated liquids. This phenomenon has been named in the literature *rectified diffusion* [32,33]. In addition, the nonlinear bubble resonances and the dynamic Blake threshold can give rise to diffusionally *stable* bubble sizes [10] This is used for example in bubble traps, where single bubbles can be captured in an acoustic standing wave field close to a pressure antinode. Bubble trap experiments have become famous for the single-bubble experiments on sonoluminescence [30]. In water, usually significant degassing was used to reach the diffusional equilibrium, but it has been demonstrated that also in saturated liquid diffusion-stable single bubbles can be achieved [28].

For a spherically and periodically oscillating bubble, the gas diffusion is governed

[9] This rule is only approximately valid, and H_i depends on both the solute and the solvent.

[10] The term *equilibrium radius* introduced above and its symbol R_0 denotes the equilibrium with respect to static pressure conditions, i. e., the resulting bubble radius if suddenly the acoustic pressure would be turned off and the bubble came to rest (therefore also *rest radius*). If we refer to the rest radius of a bubble with zero net gas diffusion, this would be denoted as *diffusional equilibrium radius*.

up to some approximation by the following equation [33–35]:

$$\frac{dR_0}{dt} = \frac{DR_GTc_0}{R_0}\left(p_0 + \frac{4\sigma}{3R_0}\right)^{-1}\left\langle\frac{R}{R_0}\right\rangle\left[\frac{c_\infty}{c_0} - \left(1 + \frac{2\sigma}{R_0p_0}\right)R_0^3\frac{\langle R/R_0\rangle}{\langle(R/R_0)^4\rangle}\right] . \quad (1)$$

Here the following symbols are used: D diffusion constant, R_G gas constant, T temperature, c_0 and c_∞ gas concentration in saturation and far from the bubble, and $\langle . \rangle$ is indicating a time average.

In an extended cavitating field, rectified gas diffusion can play a role for the life-cycle of individual bubbles and thus for the size distribution in the bubble population. This depends, among other parameters, on the gas content of the liquid, spherical shape stability of the bubbles, local pressure distribution and bubble translation. This will be discussed later.

It should also be noted that the well-known degassing of liquids by ultrasound is actually based on cavitation, and that rectified diffusion processes play here an important role.

4 Acoustic forces and bubble motion

Bubbles in acoustic fields do not only perform volume oscillations, but they also move in space. These translational motions are mainly caused by acoustic forces, the *Bjerknes forces*, that are introduced in the following. To complete the picture of bubble motion, its *virtual mass* and *viscous drag* are also briefly discussed.

4.1 Primary Bjerknes forces

It is an everyday experience that bubbles in quiet liquids rise to the surface. The reason for this is the hydrostatic pressure, which is, roughly speaking, pushing with a higher force on the bottom surface than on the top surface of the bubble. Mathematically, the net force on the bubble is yielded by summing up the forces on all surface elements, and by Gauss' theorem, we find

$$\boldsymbol{F} = \int\!\!\int p\, d\boldsymbol{S} = \int\!\!\int\!\!\int \nabla p\, dV .$$

If the pressure gradient is slowly varying over the scale of the bubble's size, we can substitute the right-hand side integral by $V\,\nabla p(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is the position of the bubble center. This approximation is usually allowed, but care has to be taken in cases where the bubbles reach wavelength extensions (like degassing processes at higher frequencies).

The hydrostatic pressure turns out to be typically negligible compared to the effects caused by the sound field: in water, gravity creates a gradient of about 10 mPa/μm, while in a pressure wave of 1 bar amplitude and frequency of some MHz we find gradients reaching hundreds of Pa/μm.

The sound pressure gradients and the bubble volumes are time varying, and the net force on a bubble depends on the time average

$$\boldsymbol{F}_{B1} = \langle V(t)\nabla p(\boldsymbol{x}_0, t)\rangle , \quad (2)$$

which is called the *primary Bjerknes force*.

The standard discussion of this average considers weak bubble oscillations and a plane standing or travelling pressure wave [4,5]. Then the results can be captured by the relative phase of pressure and bubble oscillation, which depends on the bubble size R_0 in relation to the linear resonance radius R_{res}: larger bubbles are attracted by pressure nodes, and smaller by the pressure antinodes. In propagating waves, bubbles around R_{res} feel a strong force in the wave's direction, and the force diminishes for bubbles much smaller or much larger than the linear resonance size.

In the case of strongly driven bubbles, which is typical for cavitating liquids, $\boldsymbol{F}_{B1}$ becomes more complicated because of the anharmonic bubble oscillation. In particular, the collapse time of stronger and stronger expanding bubbles shifts to later instances, and thus the primary Bjerknes force in standing waves can change its sign at high pressure zones for small bubbles. This effect was first noted in the context of SBSL [36,37], and it is very important for cavitation fields: also bubbles and bubble clusters much smaller than the linear resonance radius can get expelled from pressure antinodes, and the high pressure regions may get depleted.

Figure 10 shows for the two typical ultrasonic frequencies 20 kHz and 1 MHz the regions in $R_0 - p_a$ parameter space where the bubbles are driven away from the pressure antinode due to the primary Bjerknes force. Here, p_a is the local driving amplitude at the bubble's position and can usually be identified with a spatial position in the standing wave. In the case of a traveling wave, the changes caused by nonlinear bubble oscillations are not so dramatic, although the forces can get much larger than for linear approximation, and also parameter combinations exist where the sign reverses, i. e., the bubbles feel a primary Bjerknes force against the wave direction [38].

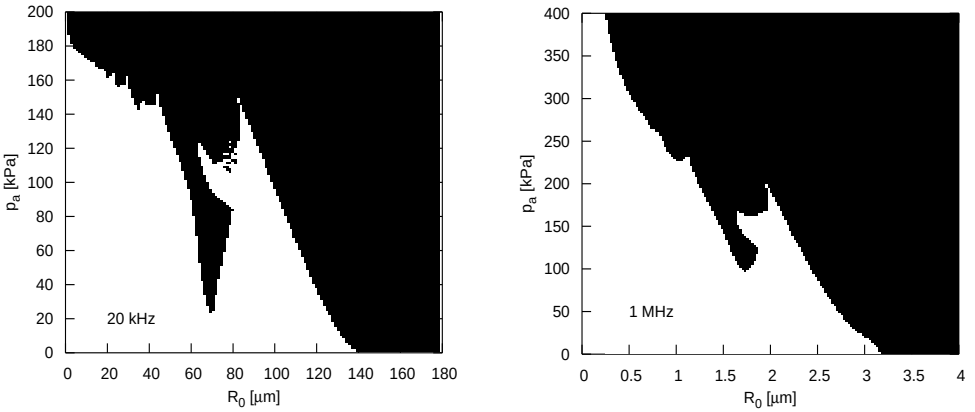


Figure 10. Sign of the primary Bjerknes force in a standing sound wave in the parameter space of bubble equilibrium radius R_0 and pressure amplitude p_a. White areas indicate a force towards the high pressure regions, dark areas a force towards the pressure nodes. Left: $f = 20\,\mathrm{kHz}$, $R_{\mathrm{res}} = 138\,\mu\mathrm{m}$; right: $f = 1\,\mathrm{MHz}$, $R_{\mathrm{res}} = 3.15\,\mu\mathrm{m}$ (water at normal conditions, $\kappa = 1$, $\sigma = 0.07275\,\mathrm{N/m}$).

A convenient notation for more general (single frequency) pressure fields is obtained by introducing a spatially varying amplitude and phase: $p(\boldsymbol{x},t) = p_a(\boldsymbol{x})\cos[\omega t + \phi(\boldsymbol{x})]$. Then the primary Bjerknes force reads

$$\begin{aligned} \boldsymbol{F}_{B1} &= -\nabla p_a(\boldsymbol{x}) \left\langle V(t)\cos\left(\omega t + \phi(\boldsymbol{x})\right)\right\rangle \\ &\quad + p_a(\boldsymbol{x})\nabla\phi(\boldsymbol{x}) \left\langle V(t)\sin\left(\omega t + \phi(\boldsymbol{x})\right)\right\rangle . \end{aligned} \tag{3}$$

Now we can associate the first term on the right-hand side as a standing wave contribution, and the second term as a travelling wave part. For the ideal cases of a plane standing (resp. travelling) wave, $p_a(\boldsymbol{x}) = \cos(\boldsymbol{k}\cdot\boldsymbol{x})$ and $\phi = const$ (resp. $\phi(\boldsymbol{x}) = -\boldsymbol{k}\cdot\boldsymbol{x}$ and $p_a = const$) for some wave vector $\boldsymbol{k}$. It is immediately clear that then the respective other terms disappear. In some situations, the sound field has the character of a decaying (or damped) travelling wave, and both terms are contributing. Then they may be counteracting and leading to spatial locations of stable force equilibrium. In the field in front of sonotrode tips, such stagnation points have been observed and modelled (Ref. [39], Sect. 6.3).

4.2 Secondary Bjerknes forces

Oscillating bubbles radiate sound, and neighboured bubbles interact via the emitted pressure field. The pressure itself adds to the acoustic driving of the adjacent bubble, and the pressure gradient leads to a net force analogous to the primary Bjerknes force in Eq. (2). It turns out that the driving coupling has often negligible effect compared to the primary (incident) pressure, but the gradient force is essential at short bubble distances. Its time average is called the *secondary Bjerknes force*, and to some approximation, it can be written [40]

$$\boldsymbol{F}_{B2}^{1,2} = -\frac{\rho}{4\pi}\,\frac{(\boldsymbol{x}_2-\boldsymbol{x}_1)}{|\boldsymbol{x}_2-\boldsymbol{x}_1|^3}\,\left\langle \dot{V}_1(t)\dot{V}_2(t)\right\rangle . \tag{4}$$

Here $\boldsymbol{F}_{B2}^{1,2}$ is the force of bubble 1 on bubble 2, and the time average is over the time derivatives of the product of both bubble volumes V_1 and V_2. It can be seen that the force decays with one over distance squared, and this is exactly like for both the electromagnetic and the gravitational force. The sign of the force, however, as well as its absolute strength, is determined by the time-averaged term. This in turn depends on several paramters like bubble size and acoustic pressure at the bubbles' locations. Different situations are therefore possible, and in the linearized case, this discussion can be reduced to the relative oscillation phases [4,5]: bubbles oscillating in phase attract each other, while in antiphase, they repel.[11] In many cases, however, the bubbles are of similar size and oscillating in similar phase, and thus the effect of the secondary Bjerknes force is frequently a mutual bubble attraction. Indeed this attracting force dominates over the primary Bjerknes force from some close

[11] It might be noted that the equality of the force of bubble 1 on bubble 2 and vice versa is an effect of the approximation of an instantaneous interaction, i. e., of infinite sound speed. Taking into account a time delay in Eq. (4) destroys the symmetry, which leads to situations where one bubble attracts the other while being itself repelled!

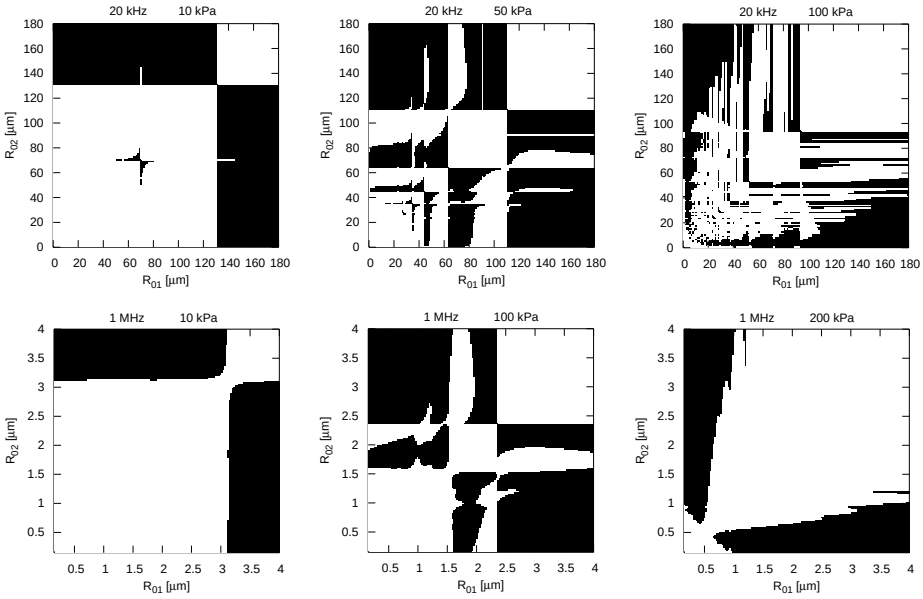


Figure 11. Sign of the secondary Bjerknes force between two bubbles of equilibrium radius R_{01} and R_{02} for fixed pressure amplitude p_a. White areas indicate an attractive force between the bubbles, and dark areas indicate mutual repulsion. Top row: $f = 20\,\mathrm{kHz}$, $R_{\mathrm{res}} = 138\,\mu\mathrm{m}$, $p_a = 10$, 50, and 200 kPa; bottom row: $f = 1\,\mathrm{MHz}$, $R_{\mathrm{res}} = 3.15\,\mu\mathrm{m}$, $p_a = 10$, 100, and 200 kPa (water under normal conditions, $\kappa = 1$, $\sigma = 0.07275\,\mathrm{N/m}$).

distance on, and bubble coalescence is typically the consequence. In Fig. 7 the subsequent approach and merging of the microbubble after ejection is in fact due to the secondary Bjerknes force. The figure also shows one reason which hinders a limitless accumulation of gas by bubble coalescence: the loss of spherical stability.

If nonlinearity is taken into account, modifications of the simple picture of the secondary Bjerknes force occur. Essentially, the absolute strength of the force can be much larger than from linear calculation, and also the size can be inverted [40]. Examples for sign distributions of the secondary Bjerknes force in the nonlinear oscillation case are shown in Fig. 11. Again 20 kHz and 1 MHz are chosen for illustration, and the driving pressures are stepwise elevated to highlight the nonlinear effects. While for 10 kPa still the classic picture with respect to the linear resonance radius occurs (a "checkerboard"), already moderately high pressure amplitudes lead to "stripes" that reflect the nonlinear resonances. The strong bubble response at the dynamic Blake threshold becomes more and more important and determines the sign of the force for small bubbles: in a way it takes over the role of the linear resonance to separate attractive and repulsive combinations of bubble sizes.

Further effects, like non-spherical distortions at close distances and coupling of oscillation and translation, have to be taken into account for a more complete picture

of the force between two oscillating bubbles. Even more, dissipative mechanisms like viscosity and heat conduction can also play an important role. However, for many aspects with respect to bubble structure formation, the spherical, nonlinear and time averaged model is a good first choice. For a much more detailed discussion of the Bjerknes forces see e. g. Ref. [41] and the references cited therein.

4.3 Added mass and viscous drag

The inertia of a moving bubble is dominated by the *virtual* or *added mass* of the liquid circulating around it, and not by its gas mass (which is typically negligible). From the assumption of potential flow, one can derive that the added mass M_a of a sphere amounts to half of the mass of the displaced liquid, i. e., $M_a = \rho V/2 = 2\pi\rho R^3/3$ [6]. The force associated with a temporal change of this inertia reads

$$\boldsymbol{F}_M = -\frac{d}{dt}\left[M_a(t)\boldsymbol{U}(t)\right] = -\frac{2}{3}\pi\rho\left[3R^2(t)\dot{R}(t)\boldsymbol{U}(t) + R^3(t)\dot{\boldsymbol{U}}(t)\right], \tag{5}$$

where $\boldsymbol{U}$ is the velocity of the bubble relative to the liquid. Note that the inertia is changed both by acceleration of the bubble centroid and by a bubble volume change caused by radial oscillation. This can lead for example to a "jerky" motion of a bubble during its collapse [42] due to conservation of (liquid flow) momentum. A direct observation of this phenomenon is shown in Fig. 12. Although the picture demonstrates that the bubble motion needs not to be smooth, sometimes an averaging approach is used to facilitate calculations. For example, the added mass is set

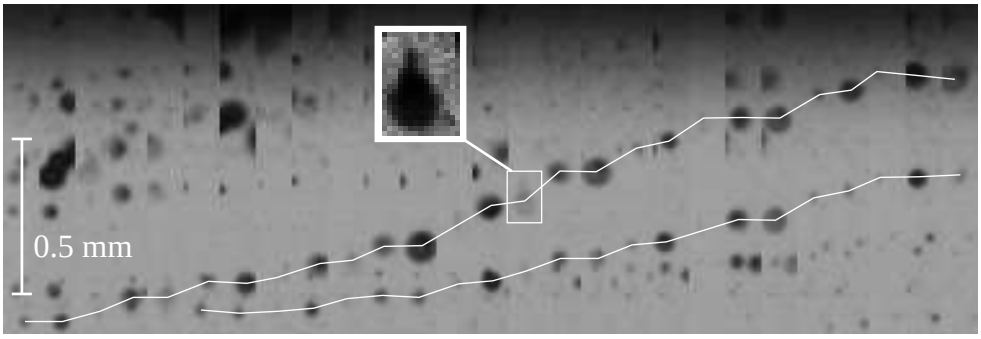


Figure 12. Pseudo-streak image from a high-speed video sequence below a sonotrode tip in water: vertical stripes from successive frames are printed next to each other, time proceeding from left to right. The acoustic frequency is 20 kHz, and the framerate 50000 fps, yielding an interframe time of 0.4 acoustic periods, which was also the exposure time. The centroid positions of two bubbles that move upwards are marked by white lines. It can clearly be observed that the bubble position frequently jumps upwards during collapse, and stops during expansion. One such jumping event is magnified in the inset, and because of the smearing due to the relatively long exposure time, the bubble shrinking simultaneous to the upward translation is revealed. Note also the relatively fast net velocity of the bubbles of 1 to 1.5 m/s.

constant to $\overline{M}_a = 2\pi\rho\overline{R}^3/3$ where the bar indicates a time average [43]. This uncouples the translational motion from the oscillation and introduces a certain error which is not always negligible [44,45].

The bubble translational motion is counteracted by a *viscous drag force*. This force can be found analytically for a spherical bubble and stationary low Reynolds number flow ($Re = 2|\boldsymbol{U}|R/\nu \ll 1$, ν being the kinematic viscosity). Depending on the supposed boundary condition at the bubble surface it reads [6,46,47]

$$\boldsymbol{F}_D^{(1)} = -4\,\pi\rho\nu R\boldsymbol{U} \qquad \text{or} \qquad \boldsymbol{F}_{\mathrm{D}}^{(2)} = -6\,\pi\rho\nu\mathrm{R}\boldsymbol{U}\,. \tag{6}$$

Here $\boldsymbol{F}_D^{(1)}$ is calculated with a zero shear stress boundary condition at the bubble surface, and $\boldsymbol{F}_D^{(2)}$ assumes a zero tangential velocity, i. e., no-slip boundary condition. While the former case reflects an "ideal" gas-liquid interface, the latter formula is valid for a "sticky" or "dirty" bubble, equivalent to a solid sphere. Indeed, surface active contaminants in the liquid may gather at the bubble interface, and in many cases experimental values from rising bubbles in non-clean liquids tend to yield $\boldsymbol{F}_D^{(1)}$ [46].

If faster bubble translations are considered, the high Reynolds number limit may be encountered, if the bubble is small enough to be still spherical and not deformed (in water up to $Re \approx 800$ [46]). In this situation the "clean" bubble boundary condition yields [46,47]

$$\boldsymbol{F}_D^{(3)} = -12\,\pi\rho\nu R\boldsymbol{U}\,, \tag{7}$$

and the "dirty" bubble case results in additional drag proportional to U^2 [46].

For oscillating bubbles, the drag becomes time dependent, and analysis gets more complicated. Some limiting cases are derived in Ref. [48]. In particular, the limit of high Re or high $(\dot{R}/U)Re$ yields for an ideal (clean) bubble interface again $\boldsymbol{F}_D^{(3)}$, but with the substitution of the constants R and $\boldsymbol{U}$ by their time-dependent counterparts $R(t)$ and $\boldsymbol{U}(t)$. Measurements of bubble motion in ultrasonic standing waves at smaller Re, but higher $(\dot{R}/U)Re$ support this result [10,49]. However, more experimental investigations, in particular in strong fields, should be undertaken to verify the theoretical approaches.

5 Bubble life cycles

If we combine all discussed aspects of single bubble dynamics, we can consider the "life" of a bubble in a given liquid and sound field. We restrict our viewpoint to translation and gas diffusion, both supposed to happen on a slower time scale than the volume oscillation. In a standing wave, the fate of a bubble is determined by its initial size R_0 and position in space $\boldsymbol{x}$, or, equivalently, the sound pressure amplitude $p_a(\boldsymbol{x})$ at the initial position. Its life cycle can be represented by a trajectory in the plane of R_0 and p_a. Main features can be seen if characteristic lines are drawn in that plane: the threshold of spherical shape instability (SI), the inversion of the primary Bjerknes force (BJ), and the rectified diffusion threshold (RD). Examples of such diagrams are shown in Fig. 13.

The general direction of a bubble trajectory in this parameter space is towards the right above the RD line, and towards the top below the BJ line. The space above

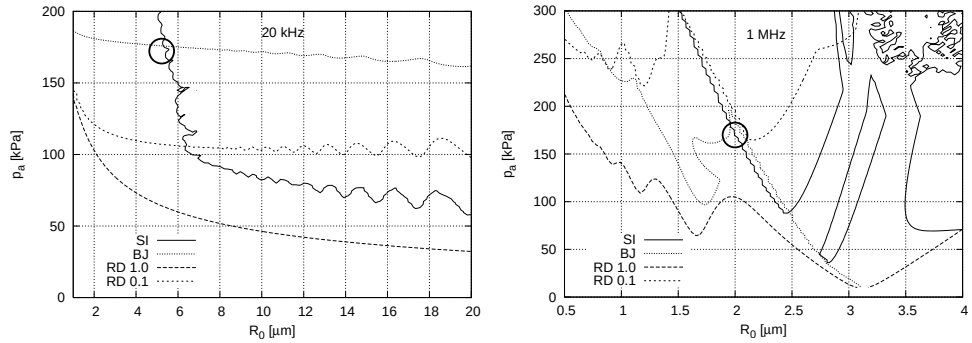


Figure 13. Phase diagrams for bubbles in a standing wave, indicating the lines of surface instability (SI) and Bjerknes force inversion (BJ). Also, the thresholds of rectified diffusion are given for water at gas saturation (RD 1.0) and degassed to 10% of saturation (RD 0.1). Left: $f = 20\,\mathrm{kHz}$, right: $f = 1\,\mathrm{MHz}$ (water at normal conditions, $\kappa = 1$). Points of accumulation are marked by a circle.

the SI line is forbidden, as bubbles there would be splitting into smaller ones soon (and thus jump to the left). From such an analysis, one can find accumulation points in the parameter plane: bubbles tend to gather at crossings of the BJ line with a negative slope part of the SI line, as long as such a point lies above the RD line. In the examples of Fig. 13, such a point is close to 175 kPa and 5 μm at 20 kHz, and 175 kPa and 2 μm at 1 MHz, respectively. The conclusion is that in standing waves of the given frequencies "typical" or "frequently found" parameters of the bubbles are those of the accumulation points.

If one considers the actual trajectories in the R_0–p_a plane, antinode pressure and wavelength have to be fixed, and added mass and viscous drag have to be taken into account as indicated in Sect. 4.3. An example is given in Fig. 14(a) where lines are plotted that represent the way of test bubbles through that plane. The shown case is interesting, because a trajectory separation along a line, similar to a water shed, can be observed. Indeed it is this water shed line which separates growing from dissolving bubbles, and not the RD threshold: bubbles moving fast enough towards a high presure region can escape from the dissolution zone [50]. This leads to the surprising result that, for the indicated parameters and irrespective of their location, almost all bubbles of rest radius larger than about 6 μm finally do *not* dissolve, but grow until surface instability sets in!

Now we briefly turn to plane travelling waves (compare also Ref. [50]). As the pressure amplitude does not vary in space, the life diagrams look simpler: for a given pressure value, it is sufficient to consider paths in the R_0–x plane, where x is the coordinate in direction of the wave vector. Examples are shown in Fig. 14(b) for 20 kHz. In the examined parameter range of small rest radii it is found that significant (macroscopic) translation appears only for growing bubbles. On the other hand, at elevated pressures the motion is rather fast, and the bubbles do not grow significantly before they covered a wavelength's distance. At intermediate pressure values, a size

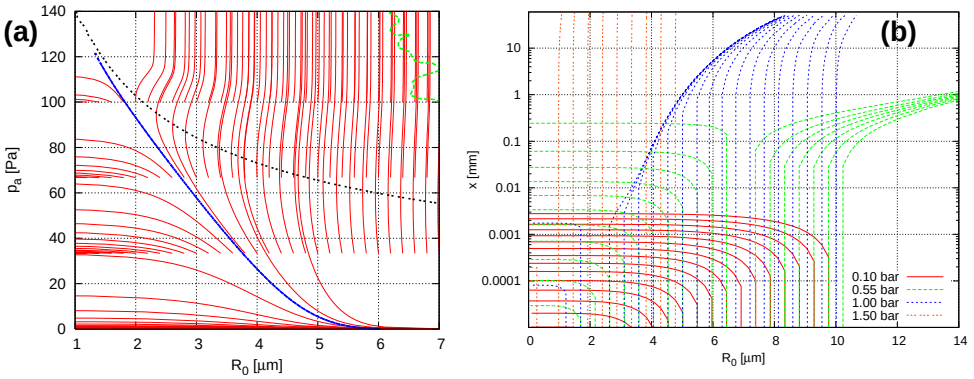


Figure 14. (a) Bubble life trajectories (red) in the plane of rest radius R_0 and pressure amplitude p_a for a harmonic plane standing wave at 20 kHz and 250 kPa antinode pressure. Trajectories generally run from bottom to the top, i. e., bubbles travel to higher values of p_a. The black dotted line is the RD threshold, and the blue line marks the effective border between dissolving and finally growing bubbles. The green, dashed line is the SI line. (b) Trajectories in the parameter plane of R_0 and x in a plane travelling wave of 20 kHz. Different colours (line styles) indicate different acoustic pressures, as given in the plot. The bubble motion is generally towards the top, i. e., to larger x (from Ref. [50]).

levelling of the growing bubbles occurs: the longer the travelled distance, the more the size distribution width of the bubble population is shrinking.

In a similar fashion, it is possible to discuss general features of bubble translation and gas diffusion in other acoustic set-ups of different frequencies or geometries. It should be kept in mind that all is based upon a single spherical bubble in a given and fixed pressure field. This approximation excludes, for instance, interaction with additional bubbles, and any influence of the bubble distribution on the sound field. At higher bubble densities, however, both effects come into play, and the life cycle inspection may turn out to be incomplete.

6 Structure formation

If we focus our interest on multibubble systems and their structures in acoustic fields, one crucial point to consider is certainly the origin of the bubbles (compare Sect. 2). However, it turns out not to be easy to assess or predict *how many* bubbles of *what size* are created *where* in the liquid. For simulations, experimentally observed data can be used as an input for this problem, and often this is still the most reasonable approach. From this, heuristic bubble generation models can be derived.

Furthermore, bubble-bubble interaction is an essential issue. With help of the secondary Bjerknes force (Sect. 4.2) and reasonable assumptions for bubble collision and merging, it is possible to "close" the theoretical description of a multi-bubble system. Between creation and destruction, each bubble follows its equations of motion, derived by the balance of forces. We add the index i to indicate the number of

the bubble and get a system of equations, where N is the total bubble number:

$$\boldsymbol{F}_M^i + \boldsymbol{F}_D^i + \boldsymbol{F}_{B1}^i + \sum_{j \neq i} \boldsymbol{F}_{B2}^{j,i} = \boldsymbol{0}\,, \quad 1 \leq i \leq N\,. \tag{8}$$

The bubble sizes are defined by the rest radii R_0^i. To describe their variation in time, a gas diffusion law like Eq. (1) can be used, augmented by rules for merging with other bubbles and splitting after surface instability.

This proposed N-body problem has some particularities, like a changing number of particles[12] and forces that depend on the absolute position in space (i. e., on the local sound pressure). It can be solved numerically, and the effort is considerably reduced if approximations of the forces are taken. In particular a temporal averaging over the oscillation period can be used, although some (typically small) errors are introduced in added mass and drag (compare Sect. 4.3).

Hinsch employed an early version of this type of particle model already in 1975 to calculate the merging process of a few bubbles [51]. Yet, the idea was further advanced only in the 1990s [43,52], probably due to the increased possibilities of computer power. In the following we present some results from recent simulations of bubble structures.

6.1 Streamer filaments

In many cases filamentary or "dendritic" arrangements of acoustic cavitation bubbles are encountered [4,5,53]. The bubbles move relatively fast along branches that can unite to form a new branch, like rivers form junctions. In a standing wave, a typical picture is that many branches arrange around a pressure antinode, and bubbles therefore stream inward the filamentary conglomerate [43]. At the center they may form a larger bubble or bubble cluster where microbubbles are shed off, sometimes visible as a "mist". The microbubbles dissolve and thus compensate for the gas arriving from the filaments.

The filamentary structure in a cubic resonator, driven at 25 kHz, has been analyzed by stereoscopic high-speed recordings in Ref. [54]. Some of the reconstructed bubble paths have afterwards been simulated by the particle approach described above. For every bubble, only the first position and velocity have been used as an input for the model, and the bubble sizes have been given a starting value of 5 μm. The pressure value at the antinode and its position have been slightly fitted within the error tolerance of a measurement. A direct comparison of experimental and calculated bubble tracks is shown in Fig. 15. The correspondence is quite reasonable, and the conclusion can be drawn that the dominant forces causing the bubble motion have been taken into account by the particle model. Still, not every detail is met perfectly by the simulation. This is probably due to variations in bubble sizes and the fine structure of the filaments: they may contain smaller bubbles, overlooked by the image processing of the experimental data, and travelling bubbles can undergo repeated split-and-merge processes. Such phenomena add "noise" to the N-body system, and because of the nonlinear bubble-bubble interaction, it is rather sensitive

[12] N is not fixed in time because of nucleation, merging, splitting, and dissolution.

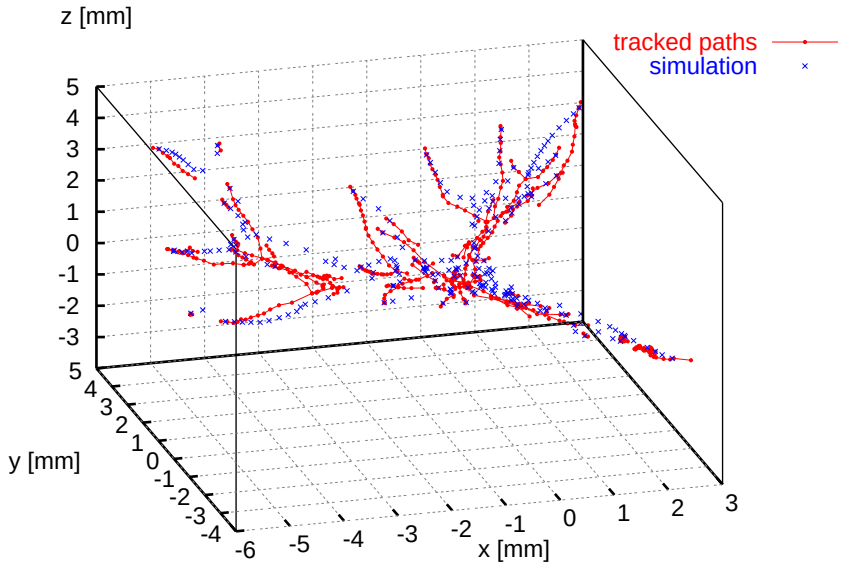


Figure 15. Comparison of experimental and simulated bubble tracks in a filamentary streamer (from Ref. [54]).

to any disturbance. Nevertheless, the simulated bubble tracks can capture essential features of the observed structures, like the gross arrangement of filaments, bubble translation velocities, or bubble collision frequencies.

6.2 Double layer structure

A bubble structure that is frequently found in intense, well developed standing waves is a double layer of dendritic filaments ("jellyfish" structure [53,55,56]). Experimental observations at 25 kHz are shown in Fig. 16(a) and (b). The two layers are arranged symmetrically to a pressure nodal plane, and each layer is situated between pressure node and antinode region. Bubbles occur from the zone in-between the layers and travel on dendrites to the layer centers. From time to time, bigger bubble clusters form in the layers and move back towards the nodal plane, dissolving on their way.

Figures 16(c) and (d) present a simulation by the particle approach. The sound field has been set to a standing wave of 25 kHz with an antinode pressure of 200 kPa, and bubble creation has been assumed to happen in ring regions close to the pressure nodal plane. It can be seen that the main aspect of the structure is captured by the model. Detailed bubble tracks have not been recorded experimentally in this case, and the validation of the simulation approach is substantially by eye. Nevertheless, the model is helpful to explain the occurence and the shape of the structure. Important aspects, though, concerning the bubble sources and the dynamics of the larger bubble clusters, still have to be clarified. Possibly the liquid flow between the layers is a further factor to be taken into account, as recent results from sonochemiluminescence measurements suggest [57].

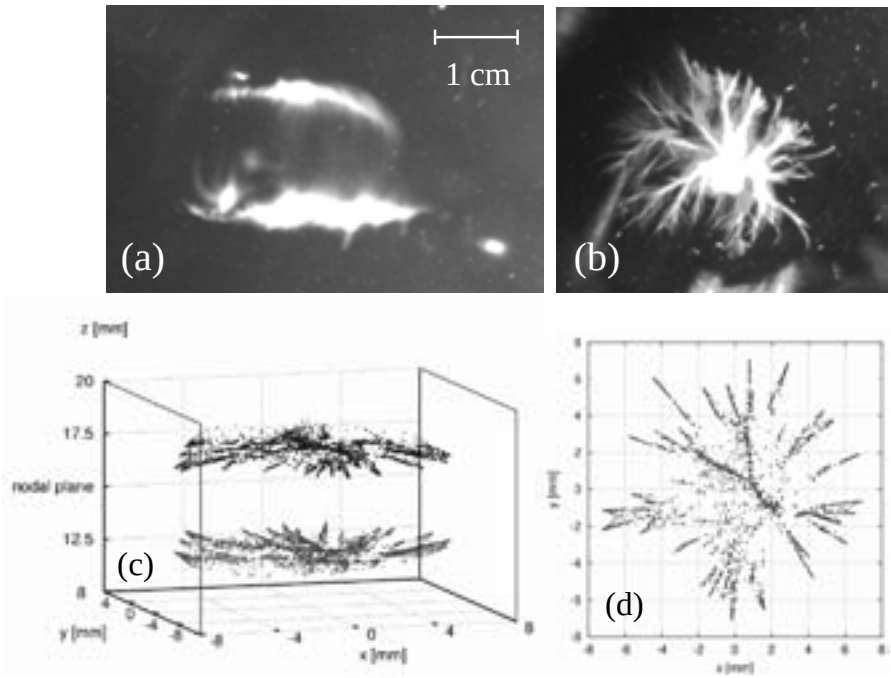


Figure 16. Double layer ("jellyfish") structure of bubbles in a standing wave. Photographs from the side (a) and from the top (b) in a 25 kHz resonator (water). Simulated bubble tracks are given in (c) and (d) (from Ref. [53]).

6.3 Cone bubble structure

As a final example we briefly discuss a conical bubble arrangement that appears at larger emitting surfaces at high intensities (Fig. 17, left). It was first reported to appear below a sonotrode (horn transducer) of larger surface area at 20 kHz [58,59], but similar bubble structures are seen, for instance, at extended submerged surface transducers [53]. Bubbles are generated directly at the emitter, and they travel rather fast away from it, in direction of the radiated wave. On their way, they form a cone, i. e., the bubble tracks meet close to a distinct point (the tip of the cone). There the bubbles stop and partly merge to form larger bubbles. Those, in turn, move further away from the sonotrode.

The pressure field forms a decaying progressive wave with a local maximum some centimetres in front of the surface, lying inside the bubbly cone and not at the cone tip [59]. The bubbles are driven to the tip point by primary Bjerknes forces, and in this case both terms on the right hand side of Eq. (3) contribute. Indeed, the tip point is for smaller bubbles a stagnation point in the sense that there both terms just cancel, i. e., the pressure amplitude gradient term and the phase gradient term are equal but of opposite sign [39]. This is not true for larger bubbles, and if smaller ones grow by rectified diffusion and merging, they sooner or later are driven again away from the tip in the forward wave direction.

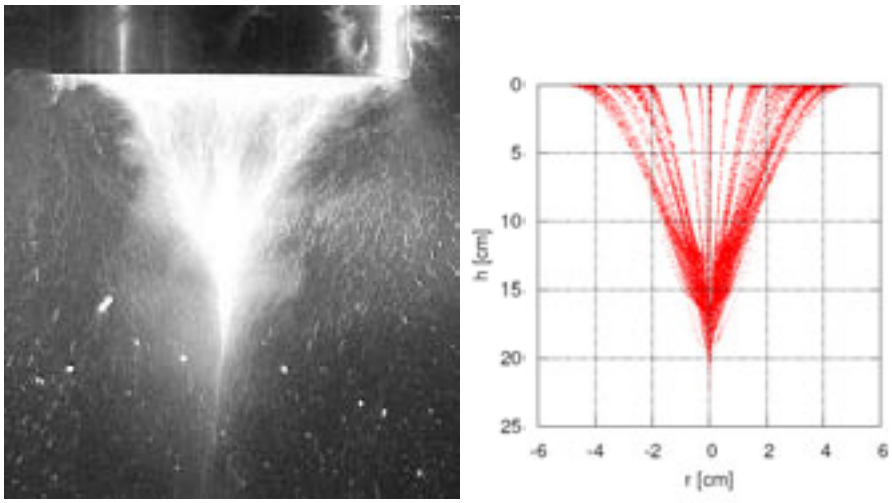


Figure 17. Left: Experimental cone bubble structure (sonotrode of 12 cm diameter, visible on top, driven at 20.5 kHz; from Ref. [59]). Right: Simulated bubble distribution (note the different scales of the axes; from Ref. [39]).

With reasonable assumptions on pressure amplitudes and bubble sizes, one can obtain a quite similar scenario and bubble structure in the particle simulation. An example is shown on the right in Fig. 17. For the calculation presented an idealized analytical pressure and phase distribution has been used, and cone extension and bubble velocities are not matching closely. Nevertheless, the structure is reproduced in principle.

7 Conclusion and perspectives

In this article, it has been outlined how macroscopic manifestations of acoustic cavitation can be approached from a single bubble point of view. The formation of cavitation fields out of individual bubbles was the motivation for a closer consideration of the forces on and the life cycles of acoustically driven bubbles. Several important factors have been discussed: the origin of bubbles, their volume oscillations and shape stability, gas diffusion, acoustic forces and bubble translation. Some experimental observations have been presented and theoretical concepts have been reviewed, partly on basis of inevitable or useful idealizations and approximations. The different aspects have been condensed into bubble life cycle diagrams that reveal a first idea of the bubble populations and conditions expected in a given set-up. Further information can be extracted from the trajectories of single bubbles in their parameter space. The next step was the treatment of multibubble fields in space and time as an N-body or interacting particle system. Although certain details of bubble nucleation and narrow interaction are still not sufficiently known, the numerical solution of the equations of bubble motion can reproduce various observed bubble

structures if suitable heuristic assumptions are included. Some examples of numerical structure reconstruction have been presented, and the results are qualitatively and partially even quantitatively satisfying.

Possible extensions of the proposed particle model approach are manifold. The most relevant physical phenomena that have been neglected in the presentation so far are probably the streaming of liquid and the modification of the initial sound field by the presence of a finite bubble density. The streaming should have a distinct influence on the (re)distribution of (passive) microbubbles and thus bubble sources. At the same time, its influence on strongly oscillating (active) bubbles is less pronounced, because their Bjerknes force induced motion is often an order of magnitude faster than the sound field induced liquid motion. The alteration of the sound field by the bubbles can happen for various reasons. They damp and scatter the sound wave and thus can cause increased dissipation and shielding. Also the impedance of the medium can change already for low bubble densities [60] which, for instance, detunes resonators. The latter effect has been investigated for a special case in Ref. [61], where the particle code has been coupled to a continuous model of sound propagation in bubbly liquid [62,63]. Indeed, a promising way to approach a better and universal simulation tool for acoustic cavitation might be a connection of (microscopic) particle simulation and (macroscopic) sound field and possibly flow field calculation on basis of continuous descriptions. First steps in this direction have already been taken, as in Ref. [61], but still many phenomena have to be investigated in more detail, and much work lies ahead until a comprehensive and satisfactory description of acoustic cavitation fields is reached.

Acknowledgements

The author thanks all actual and former members of the nonlinear dynamics and cavitation group of the DPI at Göttingen University. The friendly and stimulating atmosphere at the institute was a great help in any respect. Special thanks go to Werner Lauterborn, Thomas Kurz, and Ulrich Parlitz for continuous and strong support, fruitful discussions and pleasant joint work. I also want to thank specially Dagmar Krefting, Philipp Koch, Jürgen Appel, Jann Ohle Claussen, Till Nowak, Alexei Moussatov and Bertrand Dubus, because I used material from shared work in this article. Last but not least I thank the staff of the electrical and mechanical workshops of the institute for excellent support.

References

[1] H. G. Flynn, 'Physics of Acoustic Cavitation in Liquids', in *Physical Acoustics Vol.1B*, edited by W. P. Mason (Academic Press, London, 1964), 57–172.
[2] L. D. Rozenberg, *High-Intensity Ultrasonic Fields* (Plenum Press, New York, 1971).
[3] E. A. Neppiras, 'Acoustic Cavitation', Phys. Rep. **61**, 159 (1980).
[4] F. R. Young, *Cavitation* (McGraw-Hill, London, 1989).
[5] T. G. Leighton, *The Acoustic Bubble* (Academic Press, London, 1994).
[6] C. E. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, New York, 1995).
[7] W. Lauterborn, T. Kurz, R. Geisler, D. Kröninger, and D. Schanz, 'The single bubble

– a hot microlaboratory', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[8] L. A. Crum and D. A. Nordling, 'Velocity of transient cavities in an acoustic stationary wave', J. Acoust. Soc. Am. **52**, 294 (1972).
[9] L. Bohn, 'Schalldruckverlauf und Spektrum bei der Schwingungskavitation', Acustica **7**, 201 (1957).
[10] D. Krefting, *Untersuchung von Einzel- und Mehrblasensystemen in akustischen Resonatoren*, Dissertation, Georg-August Universität Göttingen (Göttingen, 2003).
[11] S. Yang, S. M. Dammer, N. Bremond, H. J. W. Zandvliet, E. S. Kooij, and Detlef Lohse, 'Characterization of nanobubbles on hydrophobic surfaces in water', Langmuir 23, 7072 (2007).
[12] R. D. Finch, R. Kagiwada, M. Barmatz, and I. Rudnick, 'Cavitation in Liquid Helium', Phys. Rev. **134**(6A), 1425 (1964).
[13] F. R. Gilmore, *Collapse and growth of a spherical bubble in a viscous compressible liquid*, Tech. Rep. No. 26-4, Office of Naval Research, Hydrodynamics Laboratory, California Institute of Technology, Pasadena, California (1952).
[14] J. B. Keller and M. Miksis, 'Bubble oscillations of large amplitude', J. Acoust. Soc. Am. **68**, 628 (1980).
[15] U. Parlitz, R. Mettin, S. Luther, I. Akhatov, M. Voss, and W. Lauterborn, 'Spatio-temporal dynamics of acoustic cavitation bubble clouds', Phil. Trans. R. Soc. Lond. A **357**, 313 (1999).
[16] U. Parlitz, V. Englisch, C. Scheffczyk, and W. Lauterborn, 'Bifurcation structure of bubble oscillators', J. Acoust. Soc. Am. **88**, 1061 (1990).
[17] W. Lauterborn and R. Mettin, 'Nonlinear bubble dynamics: response curves and more', in *Sonochemistry and Sonoluminescence*, edited by L. A. Crum, T. J. Mason, J. L. Reisse, and K. S. Suslick (Kluwer Academic Publishers, Dordrecht, 1999).
[18] B. E. Noltingk and E. A. Neppiras, 'Cavitation produced by ultrasonics', Proc. Phys. Soc. B **63**, 674 (1950).
[19] E. A. Neppiras and B. E. Noltingk, 'Cavitation produced by ultrasonics: theoretical conditions for the onset of cavitation', Proc. Phys. Soc. B **64**, 1032 (1951).
[20] W. Lauterborn and R. Mettin, 'Response curves of bubbles', in *Proc. of the 16th ICA/135th ASA Meeting Seattle, WA, USA* (1998), pp. 2281.
[21] I. Akhatov, N. Gumerov, C. D. Ohl, U. Parlitz, and W. Lauterborn, 'The role of surface tension in stable single-bubble sonoluminescence', Phys. Rev. Lett. **78**, 227 (1997).
[22] D. Lohse, M. P. Brenner, T. F. Dupont, S. Hilgenfeldt, and B. Johnston, 'Sonoluminescing air bubbles rectify argon', Phys. Rev. Lett. **78**, 1359 (1997).
[23] H. Lamb, *Hydrodynamics* (Cambridge University Press, Cambridge, 1932), 6th Ed.
[24] M. S. Plesset, 'On the stability of fluid flows with spherical symmetry', J. Appl. Phys. **25**, 96 (1954).
[25] H. W. Strube, 'Numerische Untersuchungen zur Stabilität nichtsphärisch schwingender Blasen', Acustica **25**, 289 (1971).
[26] T. S. Parker and L. O. Chua, *Practical Numerical Algorithms for Chaotic Systems* (Springer-Verlag, New York, 1989).
[27] D. Krefting, R. Mettin, and W. Lauterborn, 'Two-frequency driven single-bubble sonoluminescence', J. Acoust. Soc. Am. **112**, 1918 (2002).
[28] D. Krefting, R. Mettin, and W. Lauterborn, 'Single-bubble sonoluminescence in air-saturated water' Phys. Rev. Lett. **91**, 174301 (2003).
[29] D. Krefting, R. Mettin, and W. Lauterborn, 'Translationsdynamik levitierter Einzelblasen', in *Fortschritte der Akustik – DAGA 2001* (DEGA, Oldenburg, 2001), pp. 252–253.

[30] D. F. Gaitan, L. A. Crum, Ch. C. Church, and R. A. Roy, 'Sonoluminescence and bubble dynamics for a single, stable, cavitation bubble', J. Acoust. Soc. Am. **91**, 3166 (1992).
[31] P. S. Epstein and M. S. Plesset, 'On the stability of gas bubbles in liquid-gas solutions', J. Chem. Phys. **18**, 1505 (1950).
[32] D. Hsieh and M. S. Plesset, 'Theory of rectified diffusion of mass into gas bubbles', J. Acoust. Soc. Am. **33**, 206 (1961).
[33] A. Eller, 'Growth of bubbles by rectified diffusion', J. Acoust. Soc. Am. **46**, 1246 (1969).
[34] M. M. Fyrillas and A. J. Szeri, 'Dissolution or growth of soluble spherical oscillating bubbles', J. Fluid Mech. **277**, 381 (1994).
[35] O. Louisnard and F. Gomez, 'Growth by rectified diffusion of strongly acoustically forced gas bubbles in nearly saturated liquids', Phys. Rev. E **67**, 036610 (2003).
[36] I. Akhatov, R. Mettin, C. D. Ohl, U. Parlitz, and W. Lauterborn, 'Bjerknes force threshold for stable single bubble sonoluminescence', Phys. Rev. E **55**, 3747 (1997).
[37] T. J. Matula, S. M. Cordry, R. A. Roy, and L. A. Crum, 'Bjerknes force and bubble levitation under single-bubble sonoluminescence conditions', J. Acoust. Soc. Am. **102**, 1522 (1997).
[38] P. Koch, D. Krefting, T. Tervo, R. Mettin, and W. Lauterborn, 'Bubble path simulations in standing and traveling acoustic waves', in *Proc. of the 18th Int. Congress on Acoustics ICA 2004, April 4-9, 2004, Kyoto (Japan)*, pp. V-3571 (on CD-ROM).
[39] P. Koch, R. Mettin, and W. Lauterborn, 'Simulation of cavitation bubbles in travelling acoustic waves', in *Proceedings CFA/DAGA'04 Strasbourg*, edited by D. Casseraeu (DEGA, Oldenburg, 2004), pp. 919.
[40] R. Mettin, I. Akhatov, U. Parlitz, C. D. Ohl, and W. Lauterborn, 'Bjerknes forces between small cavitation bubbles in a strong acoustic field', Phys. Rev. E **56**, 2924 (1997).
[41] A. Doinikov, 'Acoustic radiation forces: classical theory and recent advances', in: *Recent Research Developments in Acoustics* (Transworld Research Network, Trivandrum, Kerala, 2003), Vol. 1, pp. 39.
[42] T. B. Benjamin and A. T. Ellis, 'The collapse of cavitation bubbles and the pressure thereby produced against solid boundaries', Phil. Trans. Roy. Soc. A **260**, 221 (1966).
[43] R. Mettin, S. Luther, C. D. Ohl, and W. Lauterborn, 'Acoustic cavitation structures and simulations by a particle model', Ultrason. Sonochem. **6**, 25 (1999).
[44] A. Reddy and A. J. Szeri, 'Coupled dynamics of translation and collapse of acoustically driven microbubbles', J. Acoust. Soc. Am. **112**, 1346 (2002).
[45] D. Krefting, J. O. Toilliez, A. J. Szeri, R. Mettin, and W. Lauterborn, 'Translation of bubbles subject to weak acoustic forcing and error in decoupling from volume oscillations', J. Acoust. Soc. Am. **120**, 670 (2006).
[46] V. G. Levich, *Physicochemical Hydrodynamics* (Prentice-Hall, Englewood Cliffs, 1962).
[47] G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge University Press, Cambridge, 1967).
[48] J. Magnaudet and D. Legendre, 'The viscous drag force on a spherical bubble with a time-dependent radius', Phys. Fluids **10**, 550 (1998).
[49] D. Krefting, R. Mettin, and W. Lauterborn, 'Kräfte in akustischen Kavitationsfeldern', in *Fortschritte der Akustik - DAGA 2002* (DEGA, Oldenburg, 2002), pp. 260.
[50] J. O. Claussen, R. Mettin, and W. Lauterborn, 'Bjerkneskräfte und gleichgerichtete Diffusion in laufenden und stehenden Wellenfeldern', in *Fortschritte der Akustik - DAGA 2007, Stuttgart* (DEGA, Berlin, 2007), pp. 131–132.
[51] K. Hinsch, 'The dynamics of bubble fields in acoustic cavitation', in *Proc. 6th Int. Symp. on Nonlinear Acoustics*, edited by V. A. Akulichev et al. (Moscow University, Moscow, 1975), pp. 26-34.

[52] R. Mettin, C.-D. Ohl, and W. Lauterborn, 'Particle approach to structure formation in acoustic cavitation', in *Sonochemistry and Sonoluminescence*, edited by L. A. Crum, T. J. Mason, J. L. Reisse, and K. S. Suslick (Kluwer Academic Publishers, Dordrecht, 1999), pp. 138–144.

[53] R. Mettin, 'Bubble structures in acoustic cavitation', in *Bubble and Particle Dynamics in Acoustic Fields: Modern Trends and Applications*, edited by A. A. Doinikov (Research Signpost, Kerala, India, 2005), pp. 1–36.

[54] J. Appel, P. Koch, R. Mettin, D. Krefting, and W. Lauterborn, 'Stereoscopic high-speed recording of bubble filaments', Ultrason. Sonochem. **11**, 39 (2004).

[55] R. Mettin, J. Appel, D. Krefting, R. Geisler, P. Koch, and W. Lauterborn, 'Bubble structures in acoustic cavitation: observation and modelling of a 'jellyfish'-streamer' in *Proc. Forum Acusticum Sevilla, Spain, 16-20 Sept. 2002*, Revista de Acustica, Vol. XXXIII, 2002, ULT-02-004-IP.

[56] R. Mettin, P. Koch, D. Krefting, and W. Lauterborn, 'Advanced observation and modeling of an acoustic cavitation structure', in *Nonlinear Acoustics at the Beginning of the 21st Century*, edited by O. V. Rudenko and O. A. Sapozhnikov (MSU, Moscow, 2002), vol. 2, pp. 1003–1006.

[57] A. Otto, R. Mettin, and W. Lauterborn, 'Raum-zeitliche Untersuchung von Sonolumineszenz und Sono-Chemolumineszenz', in *Fortschritte der Akustik - DAGA 2007, Stuttgart* (DEGA, Berlin, 2007), pp. 127–128.

[58] A. Moussatov, C. Granger, and B. Dubus, 'Cone-like bubble formation in ultrasonic cavitation field', Ultrason. Sonochem. **10**, 191 (2003).

[59] A. Moussatov, R. Mettin, C. Granger, T. Tervo, B. Dubus, and W. Lauterborn, 'Evolution of acoustic cavitation structures near larger emitting surface', in *Proceedings of the World Congress on Ultrasonics* (Société Française d'Acoustique, Paris, 2003), pp. 955-958.

[60] K. W. Commander and A. Prosperetti, 'Linear pressure waves in bubbly liquids: comparison between theory and experiments', J. Acoust. Soc. Am. **85**, 732 (1988).

[61] R. Mettin, P. Koch, W. Lauterborn, and D. Krefting, 'Modeling acoustic cavitation with bubble redistribution', in *Sixth International Symposium on Cavitation - CAV2006* (Wageningen, The Netherlands, 2006), paper no. 75.

[62] L. Van Wijngaarden, 'On the equations of motion for mixtures of liquid and gas bubbles', J. Fluid Mech. **33**, 465 (1968).

[63] R. E. Caflisch, M. J. Miksis, G. C. Papanicolaou, and Lu Ting, 'Effective equations for wave propagation in bubbly liquids', J. Fluid Mech. **153**, 259 (1997).

Oscillations, Waves and Interactions, pp. 199–216
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-08-9

Physics of stone fragmentation and new concept of wide-focus and low-pressure extracorporeal shock wave lithotripsy

Wolfgang Eisenmenger[1] **and Udo Kaatze**[2]

[1]Erstes Physikalisches Institut, Universität Stuttgart
Pfaffenwaldring 57, 70550 Stuttgart, Germany
[2]Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. In this contribution a new fragmentation mechanism by circumferential quasistatic compression using evanescent waves in a stone is described. The high efficiency of such "squeezing" process in fragmentation experiments, utilizing a self-focussing electromagnetic shock wave generator developed at the University of Stuttgart, is shown for artificial stones as well as for clinical applications. The method relies on the wide focus of the generator, with focal diameter comparable to or larger than the stone diameter. It comprises a significantly smaller peak pressure of the shock waves, with many substantial advantages for the extracorporeal shock wave lithotripsy (ESWL). A first clinical study at seven hospitals in China, in which a total of 297 patients has been treated with wide-focus and low-pressure ESWL, revealed (i) a beneficial average number of only 1532 shock pulses for successful teratment, (ii) a small mean of 1.39 sessions per patient, (iii) a high stone-free rate of 86 per cent after a follow-up of three months, and no necessity for pain medication and auxiliary measures. Additionally, with respect to the conventional narrow-focus ESWL, larger focal width involves reduced aperture and increased positional flexibility. It also enables the treatment of larger stones with a wider spatial distribution of fragments and it avoids the necessity of X-ray control during treatment, because ultrasonic positioning appears to be sufficient.

1 Introduction

Almost three per cent of the adult population is affected by kidney stones. Kidney stones are polycrystalline concrements that form from minerals and proteins in the urine. Stones come in various compositions and in sizes varying between one millimeter and some centimeters. Until about three decades ago, stones too large to pass through the urinary tract have been removed by surgery, called lithotomy. For centuries the operations bore a serious risk. In the middle of the nineteenth century, about fourteen per cent of operations resulted in the patient's death. The introduction of general anaesthesia and aseptic conditions reduced the risk of lithotomy significantly.

The successful extracorporeal disintegration of stones in the seventies of the last century, using acoustical waves or shock waves in Saarbrücken [1] and Friedrichshafen [2], has revolutionized the treatment of urinary lithiasis. Extracorporeal shock wave lithotripsy (ESWL) has received rapid acceptance, because of its noninvasive nature, high efficiency, and ease of use. Lithotripsy is the most effective non-surgical procedure for the disintegration of kidney stones ever developed. It is an alternative to surgery for seventy to eighty per cent of patients with kidney stones. Therefore, within a very few years, ESWL became the standard treatment for stones in the kidney and ureter worldwide. Stones are comminuted into small fragments which are passed with the urine during the weeks after treatment. The risk for patients to suffer major complications is distinctly lower than with surgical stone removal. The unique development of ESWL has been described in various review articles and textbooks [3–8].

Despite increasing perfection in the lithotripter generations that followed the first Dornier HM3 device, there is currently only limited agreement on the relevant stone breaking mechanisms. Some authors consider cavitation, besides shear and spalling, as the most important force for the ultimate stone fragmentation in applications of conventional lithotripters [8,9]. Cavitation is believed to cause trauma to thin-walled vessels in the kidney and in adjacent tissue as side effects, leading to short-term complications and to scar and chronic loss of tissue function [8]. Alternatively, in experiments with focal diameters on the order of the stone diameter, evidence for a squeezing mechanism has been found [10]. In addition to the question of the stone fragmentation mechanism, it has become obvious that, since the promising introduction of the Dornier HM3, progress in ESWL techniques has been unexpectedly slow [11]. The further optimization of the physical parameters of the pressure shock waves with respect to fragmentation efficiency and avoidance of side effects has been scarce and is still matter of a lively scientific debate [7,12].

In this article we summarize the results on the efficient mechanism of stone fragmentation by wide-focus and low-pressure shock waves, which one of us has obtained at the University of Stuttgart [10]. We also review clinical studies [13,14] demonstrating successful treatment of a large group of patients with kidney stones, without necessity for preventive measures, using an advanced ESWL conception. It involves a wide-focus low-pressure lithotripter that had been developed in cooperation between the Erstes Physikalisches Institut of the University of Stuttgart and the Xixin Medical Instruments Co. Ltd., Suzhou, China.

2 Lithotripter shock wave generation

Shock wave lithotripters are currently produced by more than ten companies. As an example, Fig. 1 shows the Xixin instrument, meanwhile named XX-ES lithotripter [15], as it has been used in the clinical studies which we report here. During treatment the patient rests on an examination couch which is provided with X-ray equipment and/or ultrasonic B-scan facilities for the optimum positioning. For the treatment of the patient in either dorsal or face-down position the lithotripter is equipped with a shock wave generator in an overcouch or undercouch arrangement. The shock wave

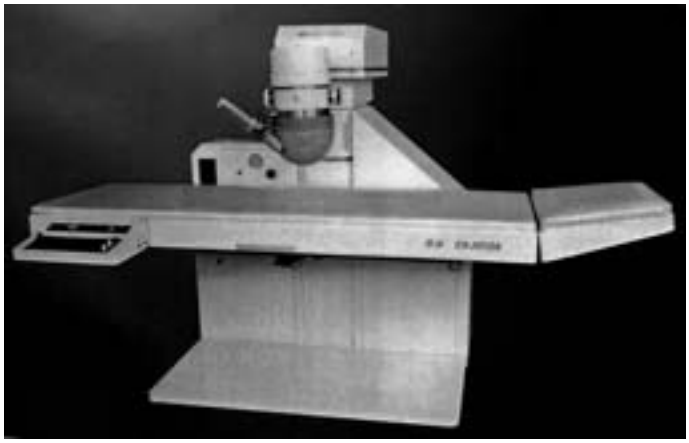


Figure 1. Lithotripter of Xixin Medical Instruments Co. Ltd., China, equipped with self-focussing electromagnetic shock wave generator from the University of Stuttgart.

is normally coupled to the patient's body via a water-filled cushion and a gel coat.

Various principles of shock wave generation are used [4]. In the electrohydraulic, or spark-gap, technology the wave is created under water by generation of an electrical discharge between two tips of an electrode (Fig. 2). The discharge produces a vaporization bubble which expands and immediately collapses, generating a high-energy pressure wave thereby. The electrode tips are located in the first focus of an ellipsoidal reflector made of suitable metal. Hence the high-energy pressure waves created in this point converge at the second focal point of the ellipse where pressure is sufficiently high for fragmentation of stones.

In the piezoelectric method a shock wave is generated by an array of some tens to thousands of piezoelectric elements stimulated with high-energy electrical pulses

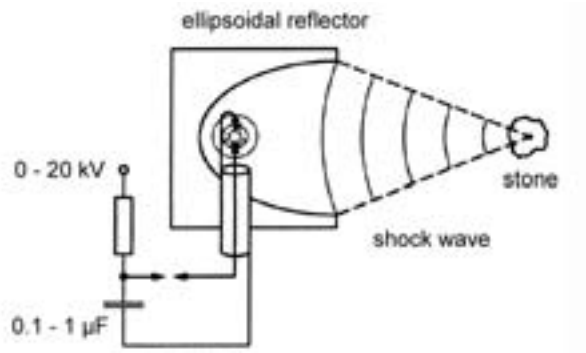


Figure 2. Sketch of electromagnetic shock wave generation with a high-voltage electrical current passing across a spark-gap electrode system that is located within a water-filled container.

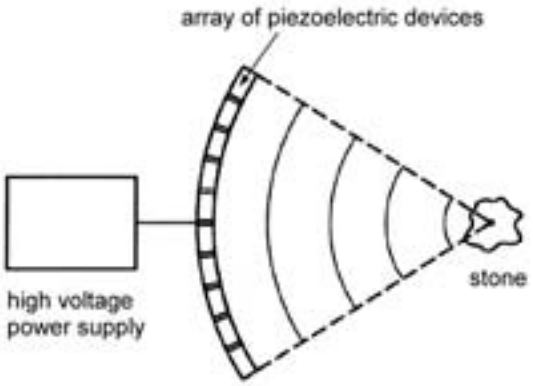


Figure 3. Principle of self-focussing piezoelectric shock wave generation.

(Fig. 3). The dome-shaped arrangement of elements in a water-filled container leads to a self-focussing shock wave.

Electromagnetic shock wave generators are based on an electromagnetic coil, positioned within a water-filled cylinder. Application of an electrical pulse results in a magnetic field which induces an opposing current in the adjacent metallic membrane, causing repulsion and pressure pulse radiation [16–18]. The membrane motions, due to nonlinearities in the wave propagation within the overlying water bath, generate shaped pressure waves. These waves are focused by either an acoustic lens (Fig. 4) or by a cylindrical reflector.

Electromagnetic generators are in favour because of their robustness, durability, reproducibility of signals, and flexibility in the choice of pulse parameters. The lens may be avoided by a self-focussing spherical shape of the coil [19] as sketched in Fig. 5. The steepening of the shock waves and the pressure profile at a series of distances from the self-focussing electromagnetic generator is shown in Fig. 6 [20].

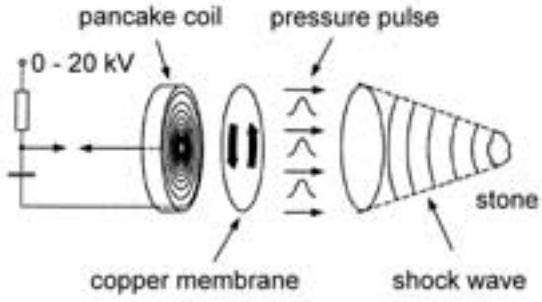


Figure 4. Electromagnetic shock wave generation using a magnetic coil to drive a metallic membrane and an acoustic lens to focus the resulting waves.

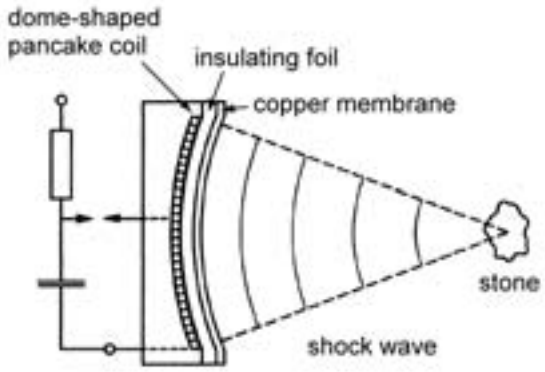


Figure 5. Self-focussing electromagnetic shock wave generation with the aid of a concavely shaped coil and membrane.

The signals have been measured with a fibre-optic probe hydrophone that basically determines optically the change in the density of the material as the pressure wave passes [21]. As revealed by the signal traces in Fig. 6, a positive pressure pulse is followed by a tensile wave. Such a profile is characteristic to almost all lithotripters. In the focus the peak amplitude of the positive pressure may vary from 10 to 120 MPa. The rise time of the shock front is in the range of nanoseconds [17,18]. The duration of the positive pressure pulse amounts to some microseconds. The amplitude of the negative pulse may reach 10 MPa. Spectral analysis of the pulses yields broad frequency bands with their centre frequencies in the range from 200 to 600 kHz.

3 Accepted stone fragmentation mechanisms

Fragmentation needs tensile stess or strain. The positive part in the pressure profile of the shock wave (Fig. 6) results in noticeable tensile stress only if it displays significant variations in space over extensions that are smaller than the stone dimensions. Pressure gradients, shear stress, as well as tensile stress and strain is then produced within the renal and urinary calculi, leading to pulverization into the desired smaller fragments. Pressure gradients are particularly high if the focus diameter of the shock waves is small as compared to the stone size. A crater-like first fragmentation erosion [22,23], as illustrated by Fig. 7, is therefore often observed with sharply focussed pressure waves.

Less sharply focussed shock waves with a pulse duration shorter than the traveling time within the stones are transmitted through the material and are reflected with pressure inversion at the rear face of the stone where the acoustical impedance changes from the large value of the solid to the smaller one of the aqueous environment. Stone material is split off [5] by the tensile stress in the reflected wave, as sketched in Fig. 8. This mechanism is known as Hopkinson effect.

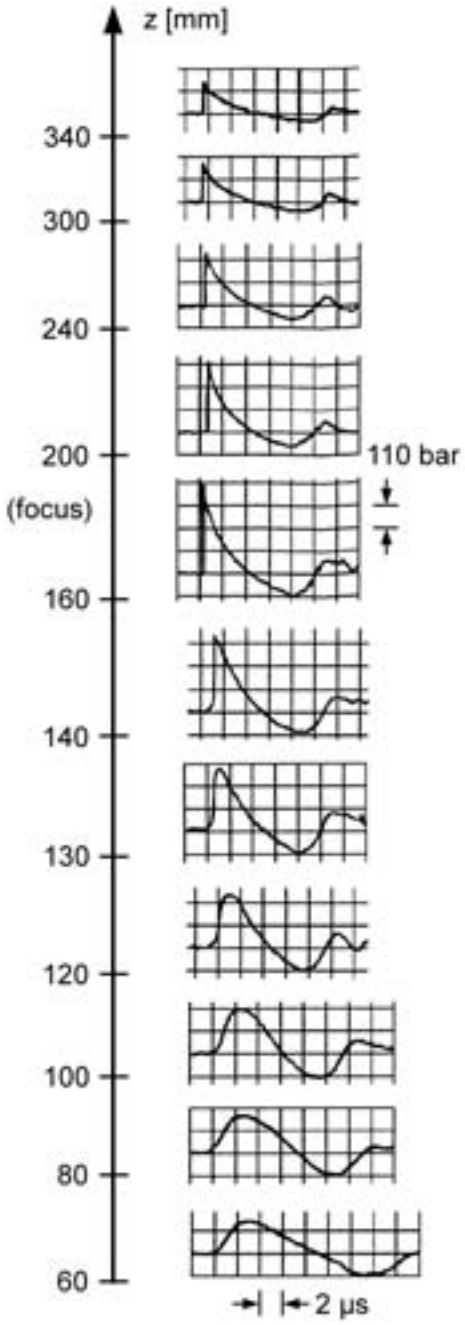


Figure 6. Steepening and focussing of shock wave profiles in dependence of the distance z from the generator [20]. The signals, produced with a self-focussing electromagnetic generator, have been recorded with the aid of a fibre-optic probe hydrophone [21].

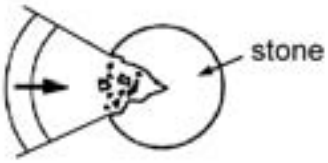


Figure 7. Scheme of crater-like fragmentation erosion of larger stones by pressure inhomogeneities of narrow-focus shock waves (-6-dB width of the focus between 2 and 6 mm).

For both above fragmentation mechanisms, which are related to the positive pressure pulse, no indications for noticeable side effects, such as tissue damage and vessel injury, exist. The situation is quite different with the action of the negative part of the shock wave profile. In addition to the direct action on the stones, the negative pressure causes cavitation in the aqueous environment and also in the liquid enclosed in microcracks [24] and cleavage interfaces of the calculi (Fig. 9). Strong evidence for significant contributions of fragmentation by cavitation exist [25].

Cavitation erosion is especially observed at the anterior and posterior side of stones [22,23]. If, however, cavitation occurs in adjacent tissues or vessels severe damage may result because, after the low pressure period of some microseconds duration, cavitation bubbles of some millimetre size are formed within some hundred microseconds (Fig. 10). These bubbles collapse rapidly, producing locally pressures of many MPa which are accompanied by high temperatures, sonoluminescence and emission of secondary shock waves [26]. Cells are ruptured upon bubble collapse [27].

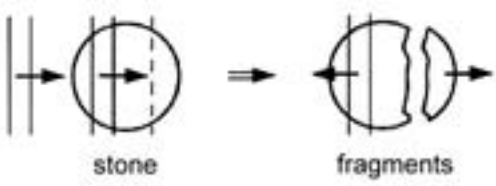


Figure 8. Principle of spalling by tensile stress in the reflected wave within the stone (Hopkinson effect).

4 Cleavage by squeezing

Crater-like stone erosions, spalling of material by tensile stress within the stone, in particular the Hopkinson effect, and cavitation are considered for a long time the dominating causes of fragmentation. Each of the mechanisms has been observed under certain conditions. The relative importance of either of the mechanisms, however, is still unclear presently. More recently it has been shown both, experimentally and theoretically [10], that binary fragmentation by quasistatic squeezing is most efficient in ESWL. Squeezing leading to stone cleavage involves shock wave focus diameters

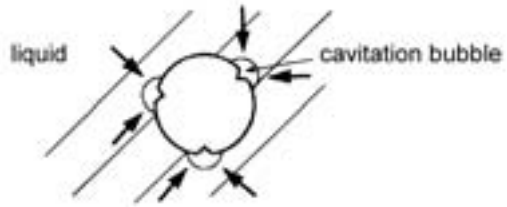


Figure 9. Erosion by cavitation at the surface of the stone.

comparable to or larger than the stone diameter. Fragmentation studies on artificial stones, using a self-focussing electromagnetic shock wave generator (Fig. 5) yielded cleavage in planes parallel to the wave vector, in experiments with positive pulse-pressure amplitudes of 20 MPa, and in planes perpendicular to the wave vector, in experiments with lower pressure (10 MPa). These observations are strikingly reproducible and are in conformity with results from other authors [24,25,28]. They are also in accordance with a squeezing model [29] as sketched in Fig. 11. The model proceeds from the part of the pressure wave that propagates outside the stone and that exerts a circular pressure to cause a compression zone inside the stone. The zone propagates with the sound velocity of the surrounding aqueous liquid which is distinctly smaller than the propagation velocities of the elastic waves within the stone. Corresponding with the pulse width in the aqueous liquid, the width of the resulting inhomogeneous pressure region amounts to 1 to 3 mm and causes tensile stress in the adjoining nonpressurised stone areas.

In the case for which the squeezing situation is depicted in Fig. 11 the situation of wave position at the centre plane of the originally globular stone is given. In Fig. 12 the first cleavage parallel to the direction of wave propagation is shown for an artificial stone of 15 mm diameter that had been exposed to seven shock wave pulses

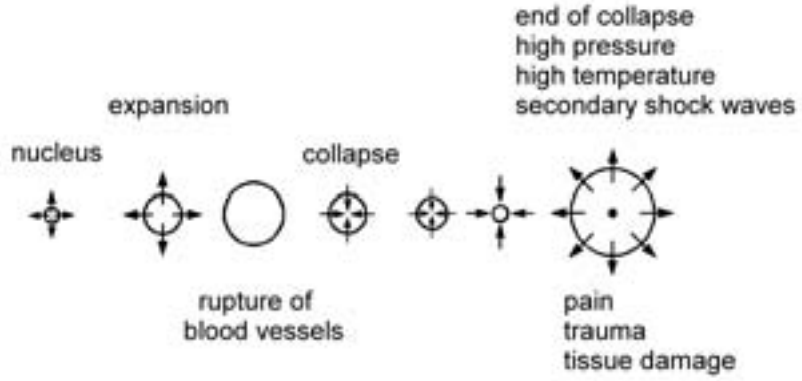


Figure 10. Cavitation bubble and its potential effects on tissue.

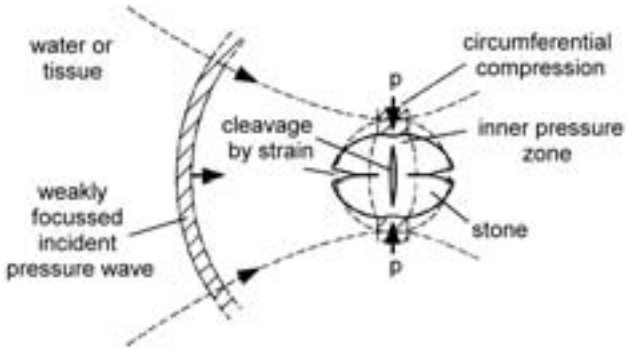


Figure 11. Orientations of cleavage planes resulting from circumferential compression (squeezing) of an originally globular stone. Within the stone the resulting strain is parallel to the direction of wave propagation. At the anterior and posterior surfaces, respectively, the resuling strain is perpendicular to that direction.

of 32.5 MPa positive amplitude, a pulse duration of 1.5 μs, and 17 mm focal diameter according to the -6 dB criterion. Fig. 13 presents the first cleavage of a similar stone in which an increased pulse pressure of 37 MPa resulted in three cleavage surfaces. With further increase of pulse pressure up to five first cleavage planes parallel to the wave vector have been observed [10].

Here we are interested in lower pressure amplitudes rather than in an immediate fragmentation into particles of 2 mm diameter or even smaller sizes. Such small fragments, that are able to pass the ureter without difficulty, are obtained from a larger number of shock wave pulses. In Fig. 14 the development of fragmentation is shown for four individual artificial stones of 15 mm diameter at varying number of pulses. The positive pressure amplitude was 25 MPa throughout, the pulse width was 1 μs and the -6-dB focal diameter was 22 mm. The series of experiments indicates an increasing number of smaller fragments with fairly narrow size distribution as the number of pulses is increased. This result is in nice accordance with a binary fragmentation mechanism [29] of large focus ESWL.

A quantitative model of the cleaving process has been developed to verify binary fragmentation by quasistatic squeezing [10]. In principle, cleaving can be described by nucleation, growth, and coalescence [12,31] of microflaws or microcracks under the repeated action of strain pulses (Fig. 15). The process proceeds until a complete crack or fragmentation interface is generated. Preexisting microflaws in kidney stones, as in many composite materials, act as nuclei for the growth of microcracks [32]. At repeated application of pressure pulses, with amplitudes just above the breaking threshold, microcracks grow and, after a number of pulses, coalesce to larger fissures. This coalescence leads to disintegration of the stone into two parts. When applying pressure shock waves of comparatively low amplitudes, coalescence of growing microcracks is caused by mechanical interaction and is controlled by stronger growth of microcracks in one single plane perpendicular to the strain direction.

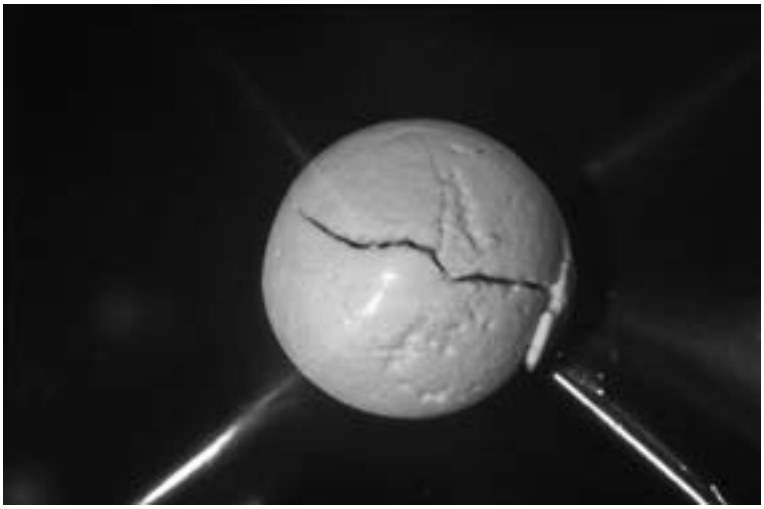


Figure 12. First cleavage parallel to the direction of wave propagation: artificial stone supplied by High Medical Technologies A.G., Switzerland; stone diameter 15 mm; 7 shock wave exposures at 32.5 MPa positive pressure amplitude, 1.5 μs pulse duration, and 17 mm focal diameter, according to the -6-dB criterion, observed parallel to the direction of wave propagation.

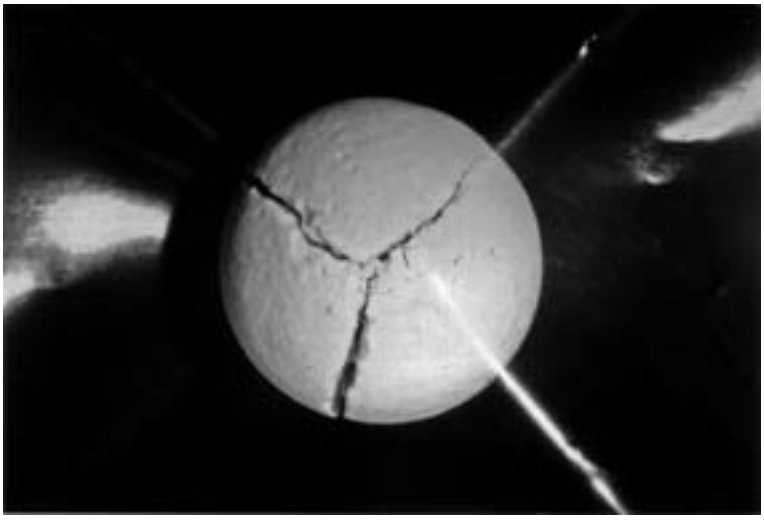


Figure 13. First cleavage of a 15 mm diameter stone as in Fig. 12 but in three cleavage surfaces. The positive pulse pressure amplitude was 37 MPa in the experiment. The direction of wave propagation was perpendicular to the figure plane.

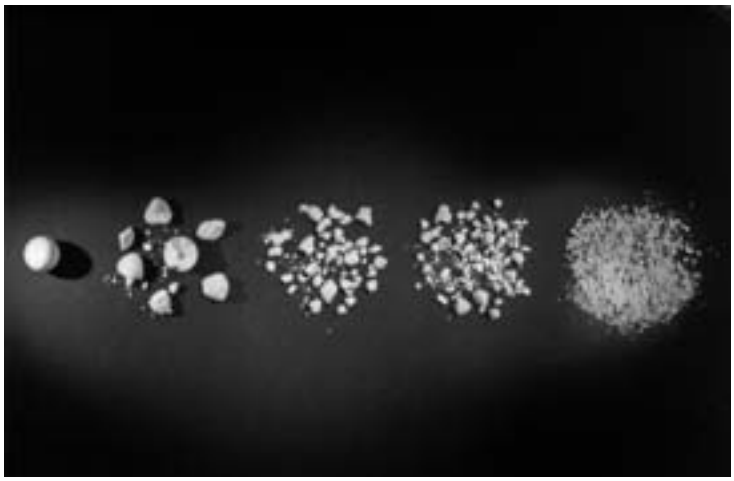


Figure 14. Results of fragmentation experiments with 15 mm diameter artificial stone as in Figs. 12 and 13, shown for different number of shock wave exposures. From left to right the figure shows first the original stone and then fragmentation after 7, 60, 120, and 500 pulses. In these experiments shock waves with positive pulse pressure of 25 MPa, pulse duration of 1 μs, and -6-dB focal diameter of 22 mm were applied. Each fragmentation result has been obtained with an extra stone. The particle size after 500 pulses was smaller than 2 mm.

The theoretical description proceeds from the observed cleavage orientations which suggest the positive part of the pressure pulse to act on the stone and its fragments by quasistatic squeezing as outlined above. It is additionally presumed that the inhomogeneous strain distribution (Fig. 11) inside the stone and the fragments does not significantly change until the lower limit of the clinically relevant size of 2 mm is reached. Equivalence of the quasistatic squeezing energy for fragmentation and the surface generation energy leads to the following expression for the number m of steps for binary fragmentation to end up with fragments of diameter d_m:

$$m = 3\ln(d_0/d_m)/\ln(2)\,, \tag{1}$$

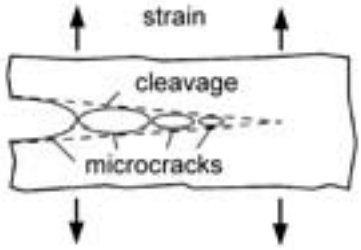


Figure 15. Scematic representation of the growth of microcracks at the repeated action of strain pulses and the resulting coalescence to a macroscopic fissure.

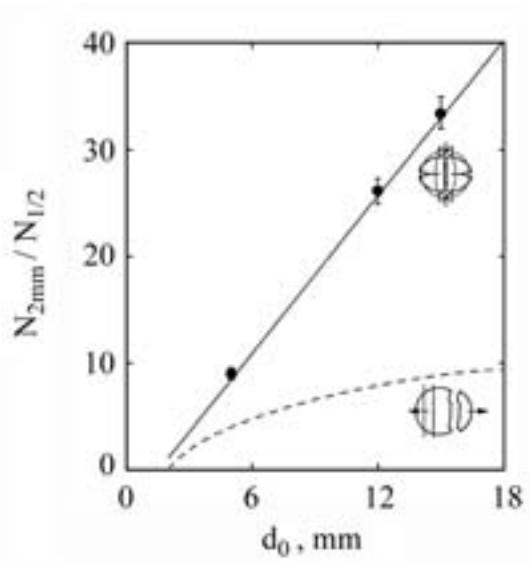


Figure 16. Ratio of the number of pulses for a 2 mm size fragmentation to that for the first cleavage displayed as a fuction of stone diameter. Figure symbols show experimemtal results for artificial stones with errors referring to measurements of 12 stones of 5 mm diameter, 10 stones of 12 mm diameter, and 7 stones of 15 mm diameter. The full line represents the theoretical relation as predicted by the quasielastic squeezing model [10]. Assuming binary fragmentation by a Hopkinson mechanism (Fig. 8) the dependence indicated by the dashed line follows.

where d_0 denotes the diameter of the original stone. In deriving this relation it has been assumed that binary fragmentation on average results in fragment volumes half the original particle volume and that the fragments can be simply considered globular particles with reduced radius. The fragmentation ratio, defined as the number of shock wave pulses needed for the production of particles with diameter d_m, divided by the number of pulses for the first cleavage, is then given by the relation [10]

$$\frac{N_{d_m}}{N_{1/2}} = \frac{2^{1/3} d_0/d_m - 1}{2^{1/3} - 1} \,. \tag{2}$$

For a final particle diameter $d_m = 2$ mm Fig. 16 shows the fragmentation ratio as a function of d_0. The full line represents the theoretical predictions based on the assumption of a squeezing mechanism (Eq. 2).

Experimental data for three sizes of artificial stones nicely agree with theory, thus confirming the idea of binary fragmentation due to quasistatic squeezing. Obviously, crater-like stone erosion, cavitation, and spalling of material by tensile stress within the stone are of minor importance at small pressure amplitudes. For the latter effect the theoretical dependence (Eq. 1) is shown for comparison by the dashed line in Fig. 16. It predicts a significantly smaller fragmentation ratio at a given stone diameter.

5 Clinical results with wide-focus low-pressure lithotripter

The above results show that binary fragmentation by squeezing is a very efficient mechanism for kidney stone disintegration. Consequently, wide-focus and low-pressure extracorporeal shock wave lithotripsy appears to be in favour when compared to conventional treatment using sharply focused pulses. The requirements in the aperture and in the placement of the generator are much lower and, in addition, the reliability to hit the stones and their fragments is considerably higher when applying wide focus waves. Most important for clinical use is the option of relatively low positive pressure and, consequently, of negative pressure amplitudes smaller than 5 MPa. Adverse reactions, such as tissue damage and pain can be avoided thereby. A short rise time of shock wave pulsesappears to be less important in applications, whereas the pulse width can be increased to a duration of 2 μs.

The conception of using an increased focus and a reduced pressure was the basis of a clincal study which has been performed in a scientific cooperation between the University of Stuttgart, the Xixin Medical Instruments Co. Ltd., and seven hospitals in China [13]. In these first studies into the reliability of the method, about three hundred patients have been treated using a lithotripter as shown in Fig. 1, equipped with the self-focussing wide-focus electromagnetic shock wave generator from the University of Stuttgart [10]. Its apperture is 120 mm and its distance of the geometrical focus is adjusted at 200 mm. At pressure amplitudes between 10 and 25 MPa in the positive pulse range, the -6-dB focal width is 1.8 μs as determined at the focus with 10 MPa positive pulse pressure. When the pressure is increased to 27.5 MPa, the -6-dB pulse duration is reduced to 1 μs in the focus. Laterally displaced from the focus, however, the -6-dB pulse length is still 1.8 μs. Figure 17 shows an example of a pressure pulse at the geometrical focus and also in the focal plane but 9 mm off axis. The pressure amplitudes depend on the generator voltage. In Table 1 the positive and negative focal peak pressure values of the generator are displayed for some generator volages. The pulse repitition rate of the generator can be adjusted in the range from 0.3 s^{-1} to 2 s^{-1}.

A detailed protocol has been kept for each patient during treatment. It involved fifty items, most of them in correspondence with a former clinical ESWL study [33], but with additional details for pain. For stone disease diagnosis as well as for stone position and for the determination of the stone size X-rays have been used throughout. The stone-free rate was checked in part by X-rays but mostly by sonography

U(kV)	P^+(MPa)	P^-(MPa)
8	11.6	-4.0
9	17.7	-4.6
10	26.1	-5.6
11	31.3	-6.4
12	33.8	-7.2

Table 1. Positive (P^+) and negative (P^-) focal peak pressures of the self-focussing electromagnetic shock wave generator at some generator voltages U.

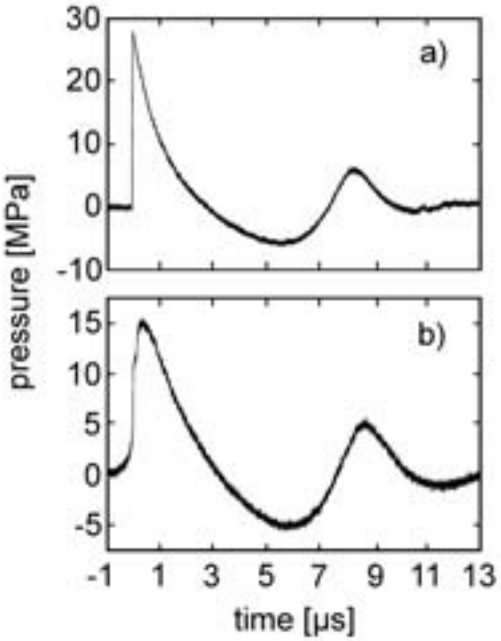


Figure 17. Shock wave pulse with positive pressure amplitude of 27.5 MPa and maximum negative presure of -6 MPa, measured with fibre-optic probe hydrophone [21] at the geometrical focus (a) and in the focal plane but 9 mm off axis (b). Corresponding with the -6-dB focal width of the pulse, the pulse pressure outside the axis is reduced to one half, but the pulse duration is close to 1.8 μs in both cases. In (b) the shock wave front has not yet developed.

after a follow-up of one week, one month, and also three months. The study included 297 patients and altogether 398 stones of which 41.6 % were smaller than 1 cm in diameter and 3.2 % were larger than 2 cm. About 32 % of the stones were located in the calyx, 30 % in the lower ureter , and 21.5 % in the upper ureter. The remaining parts were distributed across the pelvis (7 %), the middle ureter (5.5 %), and the bladder (4 %). The treatment time per session was 77 min during which the patient was exposed to a mean of 1532 shock waves with 20.8 MPa pulse pressure. The most important results of the study are summarized in Table 2 where comparison is made with clinical results obtained with narrow-focus lithotripters.

The finding that wide-focus lithotripsy required the remarkably small number of 1532 shock waves per session attracts attention, in particular as an acceptable part of only 33 % patients needed repeated ESWL and as the number of 1.39 sessions per patient was pretty small. There was no standardisation treatment procedure, but, typically, after a start with lower pressures the pulse amplitude was increased to an upper limit between 15 and 25 MPa. As a rule one thousand shock wave pulses have been applied in one session. With respect to the number of patients, as well as the number, the size distribution, and the location of stones the wide-focus

low-pressure lithotripsy study resembles a previous study by Rassweiler et al. [33] in which two narrow-focus lihotripters had been used (Table 2). Quite remarkably, in that treatment the average shock wave number was 3288 and 3457, respectively, thus more than twice the number needed with low-pressure ESWL. Additionally, the agreeable small pulse pressure of only 20.8 MPa, as a mean, applied in the wide-focus study underlines the high efficiacy of binary stone fragmentation by squeezing.

The stone-free rate of 86.3 % after three months for the treatment in all seven Chinese hospitals compares with 68 % and 70 %, respectively, in the narrow-focus lithotripsy study [33]. The figures are even more in favour of wide-focus ESWL if only the 176 patients are considered who were treated in four selected hospitals out of all seven. Just an average of 1331 shock-wave pulses was applied, with a requirement of repeated ESWL for just 22 % of the patients and a mean number of 1.24 sessions per patient. The stone-free rate was as high as 97.7 % after three months follow-up (Table 2).

The treatment time of 77 min is longer than reported for the narrow-focus lithotripters. This longer treatment time is due to the large time intervals of 3 s chosen between succeeding pressure pulses. Residual bubbles after cavitation are allowed to become progressively resorbed thereby, as has been observed by sonography. These measures reduce the contributions of bubbles to repeated cavitation events at the next pressure pulse. In the first clinical study of wide-focus low-pressure ESWL the large time interval setting aimed at a reduction of pain, traumatisation and uneasiness of the patients to the largest possible extent. In principle, the pulse repetition time can be reduced to 0.5 s.

In contrast to the narrow-focus lithotripsy study [33], for which 80 % and 77 %, respectively, of IV analgesida/sedation and 6 % general anaesthesia were reported, analgossedation was required only with 1 % of the patients treated with wide-focus ESWL (Table 2). No auxiliary measures were necessary, as compared to 38 % and 42 %, respectively, of invasive measures with conventional ESWL. Severe complications, such as the perirenal haematoma, were not reported, compared with 1.4 and 2.3 per cent of patients treated with narrow-focus lithotripsy. These results again reveal the favourable attributes of wide-focus ESWL treatment utilizing stone fragmentation by quasistatic squeezing.

Feature	Modulith SL 20	Lithostar Plus	Wide-Focus Low-Pressure	
Number of Patients	287	258	297	(176)
Number of Sessions	3288	3457	1532	(1331)
IV Analgosedation, %	80	77	1	(1)
Auxiliary Measures, %	38	42	0	(0)
Severe Complications, %	1.4	2.3	0	(0)
Stone-free Rate, %	68	70	86.2	(97.7)

Table 2. Comparison of clinical results obtained with narrow-focus lithotripters "Modulith SL20" and "Lithostar Plus" [33] as well as with the Xixin wide-focus low-pressure instrument. Figures in parentheses refer to four selected ones out of altogether seven hospitals. The stone-free rate refers to a three-months follow-up.

6 Conclusions

Based on predominant fragmentation by sqeezing, wide-focus and low-pressure extracorporeal shock wave lithotripsy has the potential of high fragmentation efficiency. A significantly reduced shock wave number is required at agreeably low pressure amplitude. At the comparatively small negative pressure range of -5 MPa and the low pulse repetition rate of $0.33\,\mathrm{s}^{-1}$, adverse effects of cavitation are less strong and do not cause severe pain. In accordance with the reduction of cavitation, as compared to conventional ESWL, only minor complications have been observed in the clinical study comprising 297 patients. Petechia and pain at the skin can be largely avoided by careful bubble-free application of the ultrasound coupling gel [10]. Missing evidence of perirenal haematoma is again in accordance with the moderate negative pressure and the thus reduced effects of cavitation. As fragmentation with wide focus results in a more homogeneous and narrow fragment-particle size distribution, auxiliary measures are necessary neither before nor after the shock wave applications.

An acceptable side effect of using a wide focus is the lower precision that is required in the localisation of stones. It is less difficult to hit the stone and its fragments during treatment. Despite of being compacted under prestress of the ureter and/or bladder entrance, middle and lower ureter stones can be easily fragmented. Also bladder stones of sizes larger than 3 cm in diameter can be sucessfully treated, provided the bladder is sufficiently filled with liquid.

Due to the many advantages offered by the wide-focus low-pressure shock-wave lithotripter, the Xixin instrument, equipped with the self-focussing electromagnetic shock wave generator of the University of Stuttgart, was given clinical approval in China.

References

[1] E. Haeussler and W. Kiefer, 'Anregung von Stoßwellen in Flüssigkeiten durch Hochgeschwindigkeits-Wassertropfen', Verhandlungen Dtsch. Phys. Gesellschaft **6**, 786 (1971).

[2] G. Hoff and A. Behrend, 'Einrichtung zum Zertrümmern von im Körper eines Lebewesens befindlichen Konkrementen', DP 2351247.2-35 (1973).

[3] F. Eisenberger, K. Miller, and J. Rassweiler, *Stone Therapy in Urology* (Thieme, Stuttgart, 1991).

[4] A. J. Coleman and J. E. Saunders, 'A Review of the Physical Properties and Biological Effects of the High Amplitude Acoustic Field used in Extracorporeal Lithotripsy', Ultrasonics **31**, 75 (1993).

[5] M. Delius, 'Medical Applications and Bioeffects of Extracorporeal Shock Waves', Shock Waves **4**, 55 (1994).

[6] C. Chaussy, F. Eisenberger, D. Jocham, and D. Wilbert, *High Energy Shock Waves in Medicine* (Thieme, Stuttgart, 1997).

[7] Delius, 'History of Shock Wave Lithotripsy', in *Proc. 15th Int. Symp. Nonlinear Acoustics, Göttingen* (Am. Inst. Phys. Press, Melville, 2000), p. 23.

[8] A. Skolarikos, G. Alivizatos, and J. de la Rosette, 'Extracorporeal Shock Wave Lithotripsy 25 Years Later', Eur. Urol. **50**, 981 (2006).

[9] J. A. Moddy, A. P. Evans, and J. E. Lingeman, *Comprehensive Urology* (Mosby Intern. Ltd., 2001), p. 623.
[10] W. Eisenmenger, 'The Mechanisms of Stone Fragmentation in ESWL', Ultrasound Med. Biol. **27**, 683 (2001).
[11] E. N. Liatsikos, 'Editorial Comment', Eur. Urol. **50**, 990 (2006).
[12] M. Lokhandwalla and B. Sturtevant, 'Fracture Mechanics Model of Stone Comminution in ESWL and Implications for Tissue Damage', Phys. Med. Biol. **45**, 1923 (2000).
[13] W. Eisenmenger, X. X. Du, C. Tang, S. Zhao, Y. Wang, F. Rong, D. Dai, M. Guan, and A. Qi, 'The First Clinical Results of "Wide-Focus and Low-Pressure" ESWL', Ultrasound Med. Biol. **28**, 769 (2002).
[14] W. Eisenmenger, in *Proc. DAGA Aachen 2003*, edited by M. Vorländer (DEGA, Oldenburg, 2003).
[15] A. P. Evan, Y. A. Pishchalnikov, J. C. Williams, J. A. McAteer, B. A. Connors, R. K. Handa, L. R. Willis, S. C. Kim, and J. E. Lingemann, 'Minimal Tissue Injury and Effective Stone Breakage in the Pig Model Using the Eisenmenger Broad Focal Zone, Low-Pressure Lithotripter', J. Urol., Suppl. **175**, 538 (2006).
[16] W. Eisenmenger, 'Eine elektromagnetische Impulsschallquelle zur Erzeugung von Druckstößen in Flüssigkeiten und Festkörpern', in *Proceedings 3rd International Congress on Acoustics*, edited by L. Cremer (Elsevier, Amsterdam, 1959), p. 326.
[17] W. Eisenmenger, 'Elektromagnetische Erzeugung von ebenen Druckstößen in Flüssigkeiten', Acustica **12**, 185 (1962).
[18] W. Eisenmenger, 'Experimentelle Bestimmung der Stoßfrontdicke aus dem akustischen Frequenzspektrum elektromagnetisch erzeugter Stoßwellen in Flüssigkeiten bei einem Stoßdruckbereich von 10 atm bis 100 atm', Acustica **14**, 187 (1964).
[19] W. Eisenmenger, Techn. Report, Deutsche Patentschrift DE 3312014 C2 (1983).
[20] J. Staudenraus, *Erzeugung und Ausbreitung freifeldfokussierter Hochenergiedruckpulse in Wasser*, Dissertation, Universität Stuttgart, Stuttgart (1991).
[21] J. Staudenraus and W. Eisenmenger, 'Fibre-Optic Probe Hydrophone for Ultrasonic and Shock-Wave Measurements in Water', Ultrasonics **31**, 267 (1993).
[22] N. Vakil, S. M. Gracewski, and E. C. Everbach, 'Relationship of Model Stone properties to Fragmentation', J. Lithotripsy Stone Dis. **3**, 304 (1991).
[23] B. Granz and G. Köhler, 'What Makes a Shock Wave Efficient in Lithotripsy?', J. Stone Dis. **4**, 123 (1992).
[24] W. Sass, M. Bräunlich, H. P. Dreyer, E. Matura, W. Folberth, H. G. Priesmeyer, and J. Seifert, 'The Mechanism of Stone Disintegration by Shock Waves', Ultrasound Med. Biol. **17**, 239 (1991).
[25] M. Delius, 'Minimal Static Excess Pressure Minimises the Effect of Extracorporeal Shock Waves on Cells and Reduces it on Gallstones', Ultrasound Med. Biol. **23**, 611 (1997).
[26] C. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, Oxford, 1995).
[27] B. Wolfrum, R. Mettin, T. Kurz, and W. Lauterborn, 'Observations of Pressure-Wave-Excited Contrast Agent Bubbles in the Vicinity of Cells', Appl. Phys. Lett. **81**, 5060 (2002).
[28] D. Holtum, *Eigenschaften und Desintegration von menschlichen Gallensteinen unter Stoßwellenwirkung*, Dissertation, Universität Stuttgart, Stuttgart (1991).
[29] W. Eisenmenger, 'Shock Wave Measuring Techniques in Liquids', in *Proc. 135th ASA Conference* (1998).
[30] S. Redner, *Statistical Models for the Fracture of Disordered Media* (Elsevier, Amsterdam, 1990).
[31] B. Sturtevant and L. Lokhandwalla, 'Biomechanical Effects of ESWL Shock Waves', in

Proc. 135th ASA Conference (1998).
[32] G. T. Camacho and M. Ortiz, 'Computational Modelling of Impact Damage in Brittle Materials', Int. J. Solids Struct. **20**, 2899 (1966).
[33] J. J. Rassweiler, K. U. Köhrmann, O. Seemann, R. Tschda, and P. M. Alken, *Kidney Stones: Medical and Surgical Management* (Lippincott-Raven Publ., Philadelphia, 1996).

Oscillations, Waves and Interactions, pp. 217–258
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-09-7

Phase transitions, material ejection, and plume dynamics in pulsed laser ablation of soft biological tissues

Alfred Vogel[1], Ingo Apitz[1], and Vasan Venugopalan[2]

[1]Institut für Biomedizinische Optik, Universität Lübeck
Peter-Monnik-Weg 4, D-23562 Lübeck, Germany
[2]Dept. of Chemical Engineering and Materials Science and
Laser Microbeam and Medical Program, Beckman Laser Institute
University of California, Irvine, California 92697, USA

1 Introduction

After the invention of the pulsed ruby laser by Maiman in 1960 it was anticipated that lasers would enable the cutting and removal of biological tissue with unprecedented precision and selectivity [1]. However, even though pulsed lasers were used within a few years for intraocular tissue *coagulation*, a clinically viable application of pulsed laser *ablation* was not reported until the early 1970's. It would take until the 1980's for lasers to be routinely used for ophthalmic dissection and ablation procedures [2,3]. In other medical sub-specialties routine laser use did not begin until the mid 1980's. The delay between the invention of the laser and its successful clinical application was largely due to a lack of understanding of the fundamental mechanisms that govern laser–tissue interactions. Now, as we approach the 50th anniversary of the invention of the laser, the understanding of laser–tissue interactions has matured and procedures that employ pulsed laser radiation are not only present in nearly every medical sub-specialty but also in various biological fields.

In the 1990's, two books on laser–tissue interactions became available and serve as a valuable resource for the field [4,5]. However, a comprehensive review of the fundamental mechanisms involved in pulsed laser ablation of tissue appeared only very recently when Vogel and Venugopalan [6] presented a conceptual framework providing mechanistic links between various ablation applications and the underlying thermodynamic and phase change processes. The present article summarizes key elements of this analysis, with focus on the kinetics of rapid phase transitions in aqueous media and their modifications by the presence of a tissue matrix. The results obtained are then used to analyse the mechanisms of material ejection in pulsed laser tissue ablation and their consequences for ablation efficiency, precision, and collateral damage. Description of the ablation plume dynamics will be an essential part of this analysis because the composition of the plume reflects the kinetics of

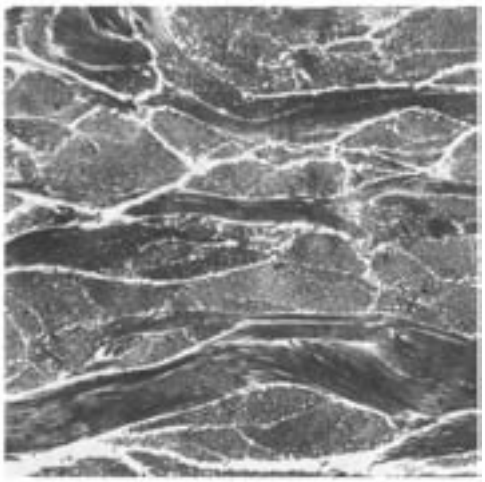


Figure 1. Transmission electron micrograph of human skin (dermis) showing collagen fibres sectioned both longitudinally and transversely. Magnification 4900×. The fibres consist of individual fibrils that are embedded in a ground substance with high water content (Reprinted with permission from Ref. [8]. Copyright 1988 Blackwell Science).

the initial phase transitions, and its rapid movement produces a recoil pressure that may result in secondary material ejection. We will confine the presentation to pulsed ablation at tissue surfaces that does not involve plasma formation. Ablation processes within transparent tissues or within cells that involve nonlinear absorption have been considered in recent reviews [6,7].

2 Tissue composition, and properties relevant for ablation

Soft biological tissues consist of cells that reside in and attach to an extracellular matrix (ECM). By mass, the composition of most soft tissues is dominated by water (55-99%) and collagen (0-35%). In "cell-continuous" tissue such as liver and epithelia, the ECM fraction is quite small and consists mostly of cell adhesion proteins. By contrast, "matrix-continuous tissues" that include the corneal stroma, dermis, cartilage, and tendon have a very small cellular fraction and are almost entirely ECM. In matrix-continuous tissues (see Fig. 1), the ECM consists largely of collagen, with the collagen content being as high as 35% [6,9]. A primary ECM function is the maintenance of the tissue's structural integrity. As a result, the ECM inhibits both tissue vaporization and material removal that represent the desired outcomes of ablation processes.

In general, the *optical absorption properties of tissue* are dominated by the absorption of proteins, DNA, melanin, hemoglobin, and water. The absorption spectra of these tissue constituents are presented in Fig. 2. In non-turbid samples, optical

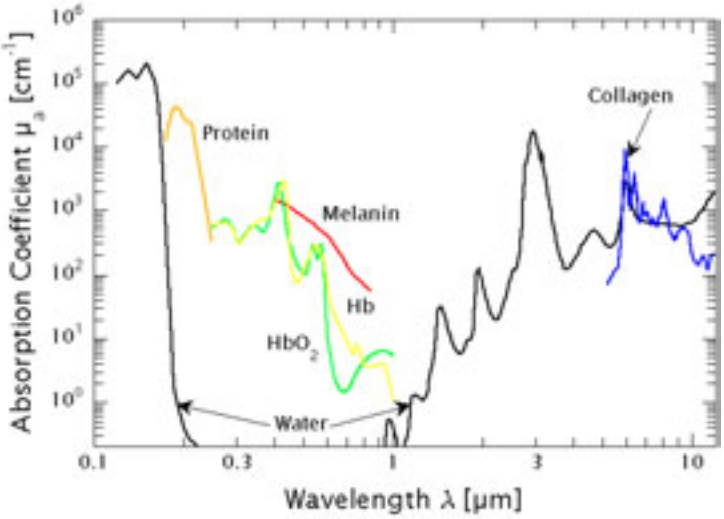


Figure 2. Optical absorption coefficients of principal tissue chromophores in the 0.1–12 μm spectral region. (Reprinted with permission from Ref. [6]. Copyright 2003 American Chemical Society)

transmission T is governed by Beer-Lambert's law according to:

$$T = \left[\frac{\Phi}{\Phi_0(1-R_s)}\right] = \exp(-\mu_a l)\,, \tag{1}$$

where R_s is the specular reflection of the sample and Φ is the radiant exposure transmitted after travel through an optical path length l in a sample with absorption coefficient $\mu_a\,[\mathrm{cm}^{-1}]$. In the absence of scattering, the reciprocal of the absorption coefficient $(1/\mu_a)$ defines the optical penetration depth δ and thus the characteristic depth of laser energy deposition.

For most laser wavelengths used only a single tissue constituent (e. g., water or collagen) absorbs the radiation. Therefore the spatial scales that characterize the distribution of these constituents are vital to understanding the potential energy transfer mechanisms between tissue constituents. The vast majority of tissue water resides in cells and in the ground substance of the ECM in which the collagen fibrils are embedded [6,10], and the spatial scale characterizing domains with different absorption properties is given by the diameter and spacing of the collagen fibrils. The fibril diameter is ≈ 30 nm in cornea and varies between 20 nm and 120 nm in dermis. The centre-to-centre spacing is very regular (≈ 65 nm) in the transparent cornea and exhibits more variations in other tissues [6,9].

Optical scattering arises from spatial variations in refractive index within tissue that are particularly strong between collagen fibrils and ground substance. Typical reduced scattering coefficients for tissues in the green are on the order of $\mu_s' = 10$–$40\,\mathrm{cm}^{-1}$ [11]. The wavelength dependence of the reduced scattering coefficient μ_s' is

well characterized by the scaling law $\mu_s' \sim \lambda^{-b}$ where $b \sim 0.5$–2 [12,13]. Optical scattering will reduce the optical penetration depth δ of light relative to the absorption depth $(1/\mu_a)$. In addition, when scattering is dominant over absorption, backscattering and total internal reflection lead to a fluence rate proximal to the tissue surface that can exceed by several times the incident irradiance [14,15]. However, ablation is typically performed at wavelengths where $\mu_a \gg \mu_s'$.

The thermal and mechanical transients generated during the pulsed laser ablation process are substantial and can result in *dynamic changes of the optical absorption properties*. Motivated by spectroscopy literature indicating that the absorption peak of water at $\lambda = 2.94\,\mu$m drops and shifts towards shorter wavelengths for increasing temperature [18], various researchers investigated the reduction in the IR absorption coefficient of tissue when heated [16,19,20]. In Fig. 3 we present data for the dependence of the absorption coefficient of water on volumetric energy ε for $\lambda = 2.94$ and $2.79\,\mu$m [16,17], and of the variation in optical penetration depth with incident radiant exposure [6]. The latter results, shown in the bottom of Fig. 3, demonstrate that for $\Phi_0 > 0.5\,\mathrm{J/cm^2}$, Er:YSGG laser radiation ($\lambda = 2.79\,\mu$m) offers better spatial confinement of the laser energy compared to Er:YAG laser radiation ($\lambda = 2.94\,\mu$m). This is opposite to the behaviour one would expect from the absorption coefficients measured at small radiant exposures (Fig. 2). Variations in optical absorption with temperature are also important for ultraviolet laser ablation. Staveteig and Walsh [21] hypothesized that absorption of UV radiation by peptide bonds is followed by heating of the surrounding water that results in a change in hydrogen bonding structure of water and thus to a red shift of the water absorption band, which at room temperature is located at 160 nm. They demonstrated that the absorption of water at $\lambda = 193\,\mathrm{nm}$ may be raised to as much as $\mu_a \approx 10^4\,\mathrm{cm^{-1}}$ at $\varepsilon = 2\,\mathrm{kJ/cm^3}$.

Phase transitions of the tissue water are strongly influenced by the *mechanical tissue properties*. There is a positive correlation between tissue strength and collagen content. Tissues that represent extremes of mechanical strength are the liver and tendon. Liver is a cell-continuous tissue with little ECM and collagen content, which results in a very low ultimate tensile strength (UTS) of 23 kPa and moderate extensibility at fracture $\sim 50\%$ [9,22]. Tendon is a matrix-continuous tissue that possesses high collagen content. This provides for high strength and stiffness with an UTS of $\gtrsim 100\,\mathrm{MPa}$ and fracture extensibility of $\sim 10\%$. Skin has similarly high collagen content (25–33%) but lower UTS ($\sim 10\,\mathrm{MPa}$) and much larger fracture extensibility of $\sim$ 30–100% because the collagen fibrils in the dermis are "wavy" and form a loose three-dimensional network [9] (see Fig. 1).

Nearly all tissue mechanical data have been acquired under "quasi-static" loading conditions in which the tissue is deformed at very slow strain rates; typically on the order of $10^{-3}\,\mathrm{s^{-1}}$. However, the processes involved in pulsed laser ablation of tissue produce extremely high strain rates; on the order of 10^5–$10^7\,\mathrm{s^{-1}}$. Studies performed to examine the effect of strain rate in the range 0.3–$170\,\mathrm{s^{-1}}$ have revealed that while the tissue strain at fracture remains roughly constant, the UTS increases in proportion to the logarithm of the strain rate [23–25]. The increase in UTS is due to the viscous dissipation between the collagen fibrils and the adjacent ground substance during the rapid deformation. It is not known whether the logarithmic dependence between strain rate and tissue UTS remains valid at the extreme strain

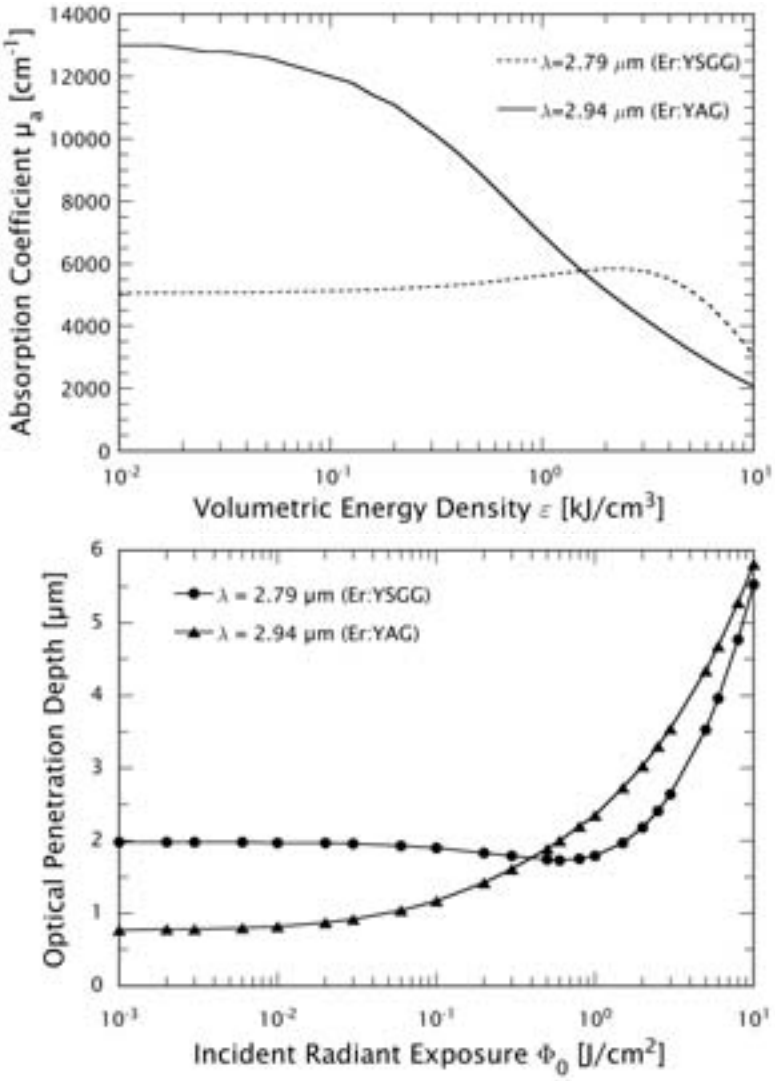


Figure 3. Graphs illustrating the effects of dynamic optical properties of a water target produced by Er:YAG ($\lambda = 2.94\,\mu$m) and Er:YSGG ($\lambda = 2.79\,\mu$m) laser irradiation; top: variation of optical absorption coefficient of water with volumetric energy density. Data compiled from Refs. [16] and [17]; bottom: variation of optical penetration depth with incident radiant exposure. Optical penetration depth is defined as the location at which the volumetric energy density drops to $1/e$ of the surface value. Note that for incident radiant exposures $F_0 > 0.4\,\mathrm{J/cm^2}$, Er:YSGG laser irradiation offers more superficial energy deposition compared to Er:YAG laser irradiation.

rates produced by pulsed laser ablation. However, the available UTS data suggest that the tissue strength under ablative conditions can be considerably higher than

that measured under "quasi-static" loading conditions.

Thermal denaturation of collagen fibrils can affect the dynamics of the ablation process. It begins when the thermal energy of the constituent molecules overcomes the weak hydrogen bonds and van der Waals interactions stabilizing the helical configuration of the α-chains in the collagen molecule [26]. The "native" triple-helical structure of the molecule is thus transformed into a "denatured" random coil structure that is associated with a loss of the banding pattern of the native collagen fibrils in TEM [27] and with shrinkage of the fibrils along their longitudinal axis [28]. However, when the collagen is embedded in tissue, shrinkage is impaired and tensile stress is developed along the fibrils due to the covalent cross-links that connect the molecules and maintain the organization of the fibrils [29,30]. Further heating denatures first the thermally-labile and then the thermally-stable covalent cross-links between the collagen molecules. This results in a stepwise disintegration of the collagen fibrils [31], a relaxation of the stresses developed during shrinkage [29,30], and finally in total mechanical failure of the fibrillar tissue structure that now appears homogeneous in TEM [26,27]. Older tissues possess a higher density of cross-links and thus require higher temperatures to undergo these transitions [28,32].

Thermal denaturation is a rate process and thus depends on both the magnitude and duration of thermal exposure [33,34]. If the heating time is reduced, considerably higher temperatures are required for denaturation. While the mechanical stability of collagen is destroyed at about 75°C when heated for several minutes [29], temperatures far in excess of 100°C are required to affect mechanical stability for thermal exposures in the nanosecond to microsecond range characteristic of pulsed laser ablation [35]. Nevertheless, given that surface temperatures approaching 400–750°C have been measured during tissue ablation using laser pulses of 100 μs duration [36], the mechanical integrity of the tissue ECM will certainly be compromised. In nanosecond ablation, temperatures in the superficial tissue layer may, even at moderate radiant exposures, be raised above 1000°C at which point the mechanical integrity of the tissue ECM is completely lost due to thermal dissociation of the constituent molecules into volatile fragments [37] (see Sect. 4.6).

Laser wavelengths especially useful for precise tissue ablation are those that exhibit very large absorption coefficients (see Fig. 2) such as the radiation of ArF excimer lasers ($\lambda = 193$ nm), Er:YSSG lasers ($\lambda = 2.79\,\mu$m), Er:YAG lasers ($\lambda = 2.94\,\mu$m), and CO_2 lasers ($\lambda = 10.6\,\mu$m). Since these wavelengths cannot well be transmitted through optical fibres, they are mainly used for ablation at tissue surfaces in air. For ablation inside the human body, XeCl excimer lasers ($\lambda = 308$ nm), thulium:YAG lasers ($\lambda = 2.0\,\mu$m) and holmium:YAG lasers ($\lambda = 2.1\,\mu$m) are frequently employed because their radiation is transmitted by low-OH quartz fibres.

3 Linear thermo-mechanical response to pulsed irradiation

In the absence of photochemical processes, the laser energy absorbed by the tissue is completely converted to a *temperature rise* before a phase transition occurs. Under adiabatic conditions, the temperature rise at a location r is related to the local

volumetric energy density $\varepsilon(r)$ by:

$$\Delta T(r) = \frac{\varepsilon(r)}{\rho c_v}\,, \tag{2}$$

where ρ is the tissue density and c_v the specific heat capacity at constant volume. The absorbed energy is redistributed by thermal diffusion [38]. In 1983, Anderson and Parrish [39] introduced the concept that spatially-confined microsurgical effects can be achieved by using laser pulse durations t_p shorter than the characteristic thermal diffusion time of the heated volume. For laser ablation, the heated volume is typically a layer with a thickness of the optical penetration depth $(1/\mu_a)$, and the characteristic thermal diffusion time t_d is given as [14]

$$t_d = \frac{1}{\kappa \mu_a^2}\,, \tag{3}$$

where κ is the thermal diffusivity. By defining a dimensionless ratio $t_d^* = (t_p/t_d)$, the thermal confinement condition can be expressed as [40,41]

$$t_d^* = \frac{t_p}{t_d} = \kappa \mu_a^2 t_p \lesssim 1\,. \tag{4}$$

Short-pulse laser irradiation of tissue not only leads to rapid heating but also to the *generation and propagation of thermoelastic stresses* [42]. The magnitude and temporal structure of these stresses are governed by the longitudinal speed of sound in the medium c_a, the laser pulse duration t_p, the depth of the heated volume $(1/\mu_a)$ and the Grüneisen coefficient Γ [42,43]. The Grüneisen coefficient is simply the internal stress per unit energy density generated when depositing energy into a target under constant volume (i. e., isochoric) conditions. This is given by the thermodynamic derivative

$$\Gamma = \left(\frac{\partial \sigma}{\partial \varepsilon}\right)_v = \frac{\beta}{\rho c_v \kappa_T}\,, \tag{5}$$

where σ is the internal stress, ε the volumetric energy density, v the specific volume, β the coefficient of thermal expansion, ρ the mass density, c_v the specific heat capacity at constant volume and κ_T the isothermal compressibility.

Thermoelastic stresses are most prominent when the laser pulse duration t_p is smaller than the characteristic time required for a stress wave to traverse the heated volume $t_m = (1/\mu_a c_a)$ [14]. This means that the stresses are confined within the heated region during the laser irradiation. By defining a dimensionless ratio $t_m^* = (t_p/t_m)$, the "stress confinement" condition can be expressed as [40,41]

$$t_m^* = \frac{t_p}{t_m} = \mu_a c_a t_p \lesssim 1\,. \tag{6}$$

For $t_m^* \ll 1$, heating of the laser-affected volume occurs under isochoric conditions and the thermoelastic stress is maximal. The peak stress σ_p is given by [42]

$$\sigma_p = A\Gamma\varepsilon_0 = A\Gamma\mu_a\Phi_0\,, \tag{7}$$

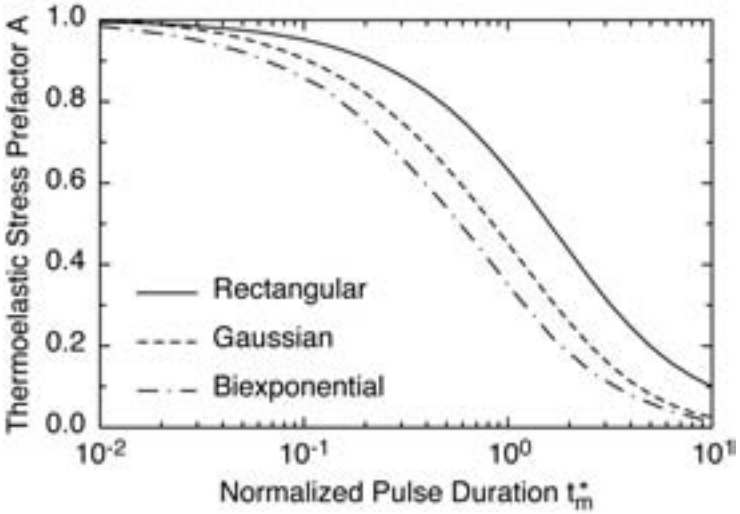


Figure 4. Variation of the thermoelastic stress prefactor A (see Eq. (7)) with pulse duration relative to the stress propagation time across the heated volume t_m^*. Results are shown for a rectangular, biexponential, and Gaussian laser pulse shape and derived from the results of Refs. [42] and [44].

where $A = 1$ and the duration of the thermoelastic stress transient t_a scales with the stress propagation time and $t_a \approx (4\text{–}6/\mu_a c_a)$. When the stress transient leaves the heated volume, the peak stress drops to $0.5\,\sigma_p$.

In the limit $t_m^* \to \infty$, where there is no stress confinement, $A \to 0$ and the duration of the stress transient approaches that of the laser pulse. The variation of A with t_m^* for different temporal laser pulse shapes is shown in Fig. 4.

While thermal expansion of a heated volume generates compressive thermoelastic stresses, subsequent propagation of these stresses results in transients that contain both compressive and tensile components. Tensile stresses arise from the reflection of the compressive stress waves at a boundary to a medium with lower acoustic impedance (tissue–air, tissue–water) or from the three-dimensional characteristics of acoustic wave propagation from a heated tissue volume of finite size [7,43,45–48]. Tensile stress wave generation originating from acoustic impedance mismatch at a tissue surface is shown in Fig. 5.

4 Thermodynamics and kinetics of phase transitions

All ablation processes involve the fracture of chemical bonds and lead to the removal of single molecules, molecular fragments, and molecular clusters. Bond fracture can also produce the formation of voids (i. e., bubbles or cracks) that facilitate the ejection of non-decomposed material fragments upon mechanical failure of the target. Vaporization, molecular fragmentation, and void formation can all be viewed as phase transitions that are accomplished via photothermal, photomechanical, and/or pho-

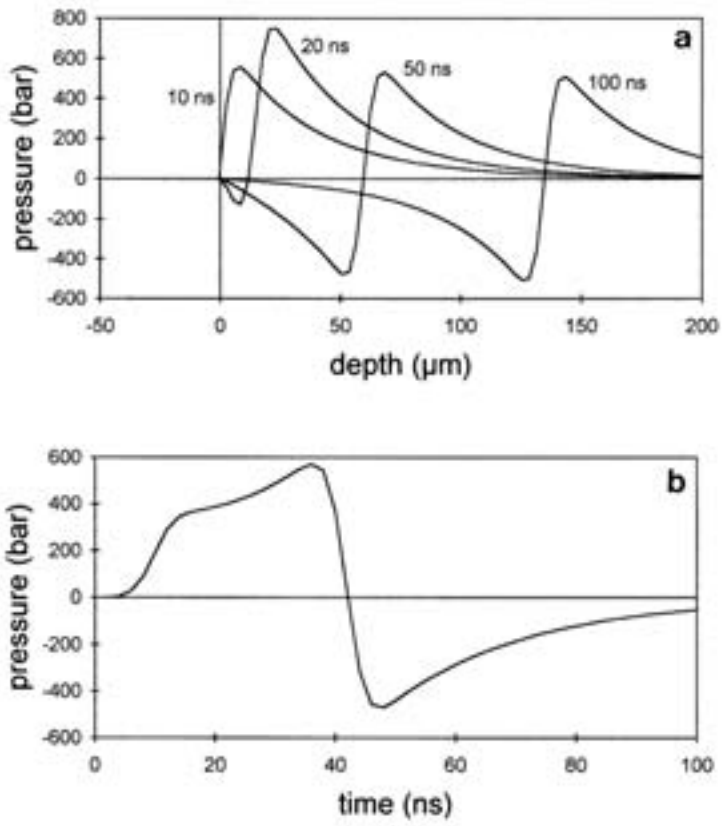


Figure 5. (a) Development of a thermoelastic stress wave in water calculated for $F_0 = 2\,\mathrm{J/cm^2}$; $\mu_a = 200\,\mathrm{cm^{-1}}$ and $t_p = 8\,\mathrm{ns}$. (b) Pressure as a function of time at a depth of $50\,\mu\mathrm{m}$. (Reprinted with permission from Ref. [45]. Copyright 1996 Springer).

tochemical mechanisms. In this article, we focus on the kinetics of phase transitions in purely photothermal ablation processes such as in IR ablation. When UV lasers are used, photochemical bond dissociation may contribute to ablation. For a detailed discussion of the role of these photochemical processes, the reader is referred to a recent review by two of the authors [6].

4.1 Phase diagrams

We use the pressure vs. temperature projection of the phase diagram for liquid and gaseous water (Fig. 6) and the pressure vs. specific volume projection of the thermodynamic phase diagram (Fig. 7) to discuss the thermodynamics of phase transitions. The solid curve A–C on Fig. 6 represents those pressure/temperature pairs where liquid and gaseous water are in equilibrium with each other and is known as the "binodal". The curve B–C–D, the "spinodal", denotes a locus of states with infinite

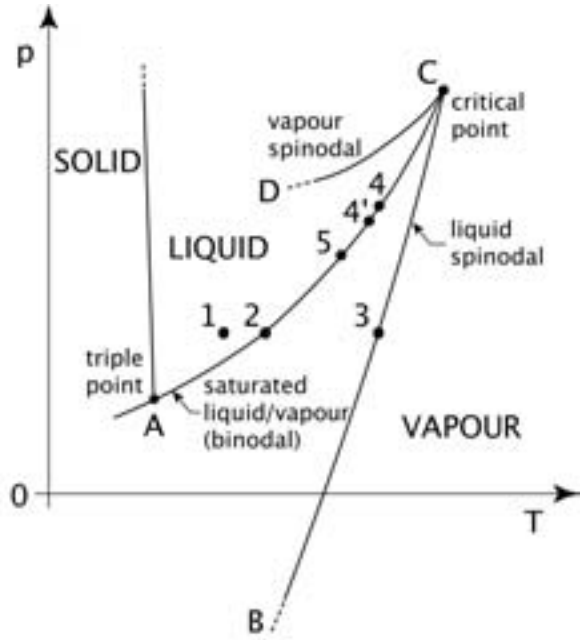


Figure 6. Pressure vs. temperature projection of the thermodynamic phase diagram including the spinodal curve. Specific states of interest are (1) ambient temperature and pressure, (2) boiling temperature under ambient conditions, (3) spinodal temperature at ambient pressure, (4) saturated conditions corresponding to the ambient spinodal temperature. The importance of points $4'$ and 5 are discussed in Sect. 5.4.

compressibility representing the intrinsic stability limit of the liquid or vapour phase $[(\partial T/\partial s)_p = 0$ and $(\partial p/\partial v)_T = 0]$. At the spinodal, the superheated liquid phase (B–C) or subcooled vapour (C–D) phase is no longer stable with respect to the random density fluctuations that are present in all materials at non-zero temperatures. The region between segment A–C of the binodal and B–C of the spinodal represents metastable states of the superheated liquid for which the presence of a free surface or bubble nucleus is required for vaporization. The binodal and spinodal curves intersect at the critical point C above which no thermodynamic distinction can be made between liquid and vapour phases. For water, the critical point is located at $T_c = 374.14\,^{\circ}\mathrm{C}$ and $p_c = 22.09\,\mathrm{MPa}$.

Liquid, vapour, and mixed phase regions are clearly demarcated in the p–v diagram of Fig. 7. The binodal encompasses the mixed phase region that specifies the range of specific volumes in which liquid and gaseous phases coexist for a given pressure and temperature. The dashed curve provides the spinodal where the segment B–C represents the stability limit of superheated liquid and segment C–D represents the stability limit of subcooled vapour.

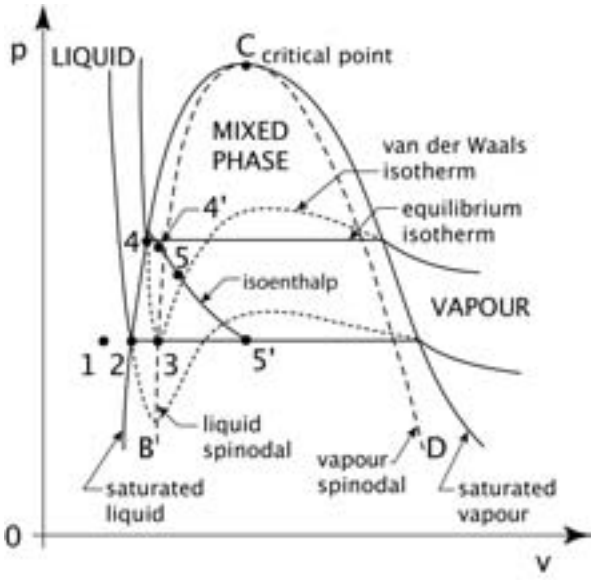


Figure 7. Pressure vs. specific volume projection of the thermodynamic phase diagram including the spinodal curve and equilibrium and van der Waals isotherms. Points 1–4 correspond to those shown in Fig. 11. The importance of points 4′, 5, 5′ and the isoenthalp is discussed in Sect. 5.4.

4.2 Surface vaporization

Equilibrium vaporization at a liquid–vapour interface is associated with an increase of specific volume at constant temperature that requires the latent heat of vaporization. This process can occur at any single location along the binodal in the p–T diagram of Fig. 6. In the p–v diagram in Fig. 7, surface vaporization is represented by a path following an equilibrium isotherm that connects a state of saturated liquid with a state of saturated vapour. Thus vaporization does not occur at a pre-determined temperature and theoretical models that adopt a fixed "vaporization temperature" [49] violate the basic physics of the process [50]. The actual surface temperature is dictated by the rate of equilibrium vaporization that balances the irradiance supplied to the system.

However, during laser ablation, rates of vapour formation in excess of that predicted by equilibrium vaporization are often achieved. This occurs because the increased equilibrium vapour pressure corresponding to the increased temperature of the liquid surface is not established instantaneously. This results in an increased mass flux of vapour transported into the surroundings known as non-equilibrium interphase mass transfer and can be estimated using arguments from the kinetic theory of gases [51,52]. Nevertheless, even this increased non-equilibrium interphase mass transfer provides a very small ablation rate. *Efficient* ablation is achieved at laser irradiances that deposit energy at a rate that cannot be balanced by surface vaporization processes alone. This results in material removal via *volumetric* processes.

4.3 Normal boiling

Normal boiling refers to a volumetric process that forms vapour at a thermodynamic state on the binodal as indicated by point 2 in Fig. 6. Thus, for a given pressure, the binodal defines the corresponding "boiling temperature". For water at atmospheric pressure this temperature is 100°C. However for ablation processes with a high rate of mass removal, the boiling temperature is increased significantly because the recoil produces an increase in pressure both at the target surface and within its bulk.

Vapour formation in normal boiling relies on the presence of pre-existing nuclei of vapour or dissolved gas within the liquid to catalyze the nucleation and growth of vapour bubbles. The transition from saturated liquid to saturated vapour occurs within a finite layer of mixed phase. The thickness of this "vapour–liquid" layer is comparable to the optical penetration depth of the incident radiation and its composition varies from that of saturated liquid at its base to saturated vapour at the target surface [50,53]. As a result, the surface temperature is fixed at the saturation conditions corresponding to the pressure at the target surface and there is no temperature gradient within the vapour–liquid layer.

It is important to note that once a normal boiling process is established, the presence of volumetric energy densities corresponding to temperatures slightly higher than the saturation temperature results in the growth of vapour bubbles. Therefore normal boiling processes always involve *partial vaporization* of a liquid volume through the growth of vapour bubbles. Thus the concept frequently found in biomedical ablation papers that vaporization only occurs once the entire latent heat of vaporization is deposited is not correct.

Nevertheless, normal boiling plays a negligible role for pulsed laser ablation for two reasons. First, the density of heterogeneous bubble nucleation sites is likely insufficient to provide a boiling process sufficiently vigorous to balance the high rates of energy deposition achieved in most pulsed laser ablation processes [41,50]. Second, the high rates of energy deposition can be balanced only if the bubbles move to the target surface on a time scale set by the propagation velocity of the ablation front. Miotello and Kelly [54] showed that this is not possible when irradiating pure water with nanosecond laser pulses and is possible for microsecond pulses only for radiant exposures proximal to the ablation threshold. In tissue this is even less likely because the mobility of vapour bubbles is further inhibited by the presence of the ECM.

4.4 Phase explosion and explosive boiling

When the rate of volumetric energy deposition provided by laser radiation is more rapid than the rate of energy consumed by vaporization and normal boiling, the tissue water is driven to a metastable superheated state. The superheated liquid is metastable until the spinodal temperature is reached. The spinodal limit is defined by line B–C in Fig. 6 that represents the locus of states with infinite compressibility $[(\partial p/\partial v)_T = 0]$. At the spinodal limit, the superheated liquid undergoes "spinodal decomposition"; a spontaneous process by which a thermodynamically unstable liquid relaxes towards equilibrium by "phase separation" into a mixture of saturated vapour and saturated liquid [55–57]. The spinodal temperature of water at atmo-

spheric pressure is $\approx 305°C$ with the corresponding equilibrium saturation vapour pressure of 9.2 MPa. Thus spinodal decomposition under atmospheric conditions involves an impressive pressure rise resulting in the violent emission of saturated liquid droplets by the expanding vapour.

For the phase diagram shown in Fig. 6, the heating phase corresponds to the path $1 \rightarrow 3$, and the spinodal decomposition will initially result in a nearly isochoric transition from point 3 on the spinodal to point $4'$ in the mixed phase region possessing the same enthalpy. For pure water, the subsequent explosive expansion of this mixture will transition through a series of thermodynamic states that follows the curve of constant enthalpy (isoenthalp) as shown in Fig. 7 until the mixture reaches atmospheric pressure at point $5'$. During the expansion $4' \rightarrow 5'$, the temperature of the mixture drops to 100°C, and about half of the liquid is transformed into vapour. The vapour fraction ($\sim 49.6\%$) is given by the energy density necessary to heat water from room temperature to the spinodal limit (1.27 kJ/g) as compared to the sum of the sensible and latent enthalpy of vaporization. The remaining saturated liquid is ejected in the form of droplets.

To provide a complete description of the phase transformation process as the liquid is heated to the spinodal limit, one must also consider the potential contribution of homogeneous nucleation [41,54,58,59]. Homogeneous nucleation refers to the spontaneous formation of vapour inclusions within the bulk liquid that arise solely from thermodynamic fluctuations and is not catalyzed by the presence of impurities or dissolved gas. While the formation of such vapour "nuclei" is spontaneous, their growth is not ensured and depends strongly on superheat temperature.

In classical nucleation theory, the driving force for growth of vapour nuclei is supplied by the difference in chemical potential between the superheated liquid outside the bubble and the vapour inside the bubble. This driving force is necessary to overcome the free energy barrier posed by the surface tension separating the vapour from the liquid [57]. The chemical potential difference between the superheated liquid and vapour scales with the bubble volume (i. e., r^3) while the contribution from surface tension scales with the bubble surface area (i. e., r^2). As a result, small vapour nuclei that form due to thermodynamic fluctuations spontaneously collapse while larger vapour nuclei will grow. The Gibbs free energy ΔG that describes the thermodynamics of bubble formation is given by:

$$\Delta G = \frac{4\pi r^3}{3}(\mu_v - \mu_l) + 4\pi r^2 \sigma \,, \tag{8}$$

where μ_v and μ_l are the chemical potentials of the vapour and liquid state, respectively, r is the size of the vapour nuclei, and σ is the surface tension of the surrounding liquid [57,60]. Nuclei grow if they are larger than a critical radius r_{cr}. Figure 8 shows the dependence of r_{cr} on superheat temperature for water. Note that while r_{cr} strongly decreases as the superheat temperature increases, it remains finite even at the spinodal temperature. Thus nucleation remains an activated process with a finite free energy barrier [55]. The strong reduction of Δr_{cr} results in a dramatic rise in the nucleation rate J with temperature that attains a large, but finite, value at the spinodal temperature as shown in Fig. 9. The energy barrier that must be overcome for the conversion from the liquid to vapour phase disappears only when surface

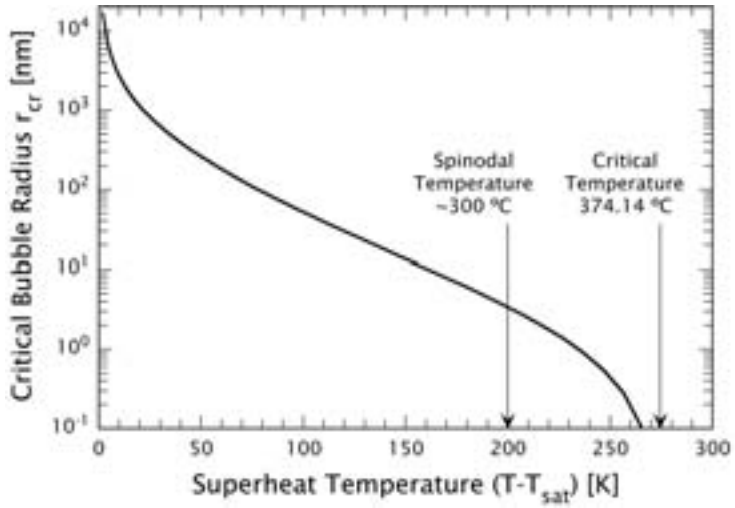


Figure 8. Variation of the critical bubble radius required for spontaneous vapour bubble growth with superheat temperature. Note that the critical bubble radius goes to zero at the critical temperature.

tension disappears and this occurs at the critical point. To account for the influence of statistical fluctuations on the stability limit, Kiselev has introduced the concept of a *kinetic spinodal*. The kinetic spinodal is defined as the locus of thermodynamic states where the time for spontaneous formation (driven by superheat temperature) of vapour nuclei becomes smaller than the characteristic time for their decay to local equilibrium (driven by surface tension) [61]. The superheat temperatures defined by the kinetic spinodal are much lower than the critical temperature and slightly lower than the classical spinodal and represent the physical limits of the metastable liquid states that can be achieved prior to spinodal decomposition.

Thus, in general, the transformation of superheated (metastable) liquid to an equilibrium state of mixed phase may involve both bubble nucleation (large density fluctuations extending over a small spatial extent) and spinodal decomposition (small density fluctuations extending over a large spatial extent). We refer to the collective phase transition process as a *phase explosion*. A more detailed consideration of nucleation theory and spinodal decomposition as it relates phase transitions and tissue ablation can be found in our earlier review [6].

Thus far we have focused on processes tracing a path indicated by $1 \rightarrow 3 \rightarrow 4' \rightarrow 5 \rightarrow 5'$ in Figs. 6 and 7. This path corresponds to the extreme case in which no vapour nuclei are present in the liquid. When the heating occurs very rapidly at high radiant exposures, the liquid experiences a recoil pressure from surface vaporization/interphase mass transfer that can be substantial due to the non-equilibrium conditions produced during the beginning of the laser pulse. Thus spinodal conditions are reached at an elevated pressure somewhere between point 3 and the critical point C. Because the resulting phase explosion occurs at elevated temperature and

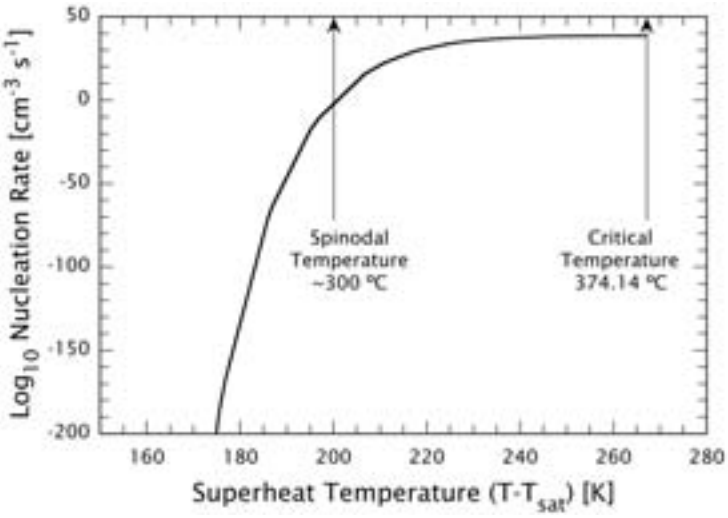


Figure 9. Variation of vapour bubble nucleation rate with superheat temperature.

pressure, the pressure jump associated with the phase separation is less severe. The elevated temperature corresponds to a higher volumetric energy density of the superheated liquid, and therefore more than half of the liquid will be transformed into vapour during the phase separation process.

When vapour nuclei are present in the liquid and the heating occurs on a time scale such that a significant fraction of the incident laser energy (but not the entire energy flux) contributes to the growth of heterogeneous and homogeneous vapour nuclei, the resulting phase transition process again follows a path that is intermediate between normal boiling and the path $1 \rightarrow 3 \rightarrow 4$. Spinodal conditions in the superheated liquid are again reached at a location between point 3 and the critical point C followed by phase separation. In this case however, the pressure rise is due to vapour formation at the nucleating centres rather than recoil from vapour leaving the target surface. Such intermediate processes are termed *"explosive boiling"*. In general, the energy necessary to reach spinodal conditions is higher for explosive boiling than for the case of phase explosion with surface vaporization. The amount of vapour formation is greater due to contributions from both the growing nucleation centres and the phase separation.

Both phase explosion and explosive boiling are volumetric processes in which a portion of the target material is ejected in the liquid phase and the latent heat of vaporization is not supplied to the entire ablated mass. As a result, the ablation efficiency (mass removed by a given amount of laser energy) is higher for these processes as compared to surface vaporization and normal boiling where all material is removed in the vapour phase.

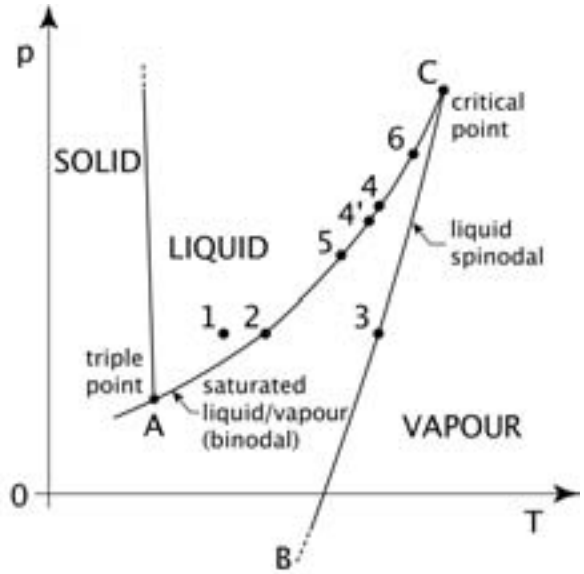


Figure 10. Path taken through the p vs. T projection of the thermodynamic phase diagram for confined boiling ($1 \rightarrow 2 \rightarrow 4 \rightarrow 6$) and for tissue ablation involving a phase explosion ($1 \rightarrow 2 \rightarrow 3 \rightarrow 4' \rightarrow 5 \rightarrow 6$). The actual path followed depends on the rate of energy deposition, number density of heterogeneous nuclei, and the mechanical strength of the tissue matrix relative to the saturation vapour pressure corresponding to the ambient spinodal temperature.

4.5 Effects of the tissue ECM on the phase transitions

In pulsed laser ablation of tissues, the phase transition processes are affected by the presence of the ECM. For boiling processes within tissue, the vapour pressure necessary to drive bubble growth must not only overcome surface tension but also the elastic restoring forces provided by the tissue matrix [62]. Therefore, bubble growth in tissue requires a higher internal pressure than in pure liquids, and the elevated pressure is coupled to an increase in the boiling (saturation) temperature. The pressure increase that develops during the boiling process continues until it exceeds the ultimate tensile strength of the ECM and results in explosive tissue ablation [62]. We term this process "confined boiling". On the p–T phase diagram in Fig. 10, the confined boiling process corresponds to a path $1 \rightarrow 2 \rightarrow 6$, where the $2 \rightarrow 6$ transition is coincident with the binodal and terminates where the saturated vapour pressure equals the ultimate tensile strength of the tissue. In the presence of a tissue matrix, explosive material ejection will thus occur regardless of the rate of energy deposition. Thus, it is not surprising that explosive material ejection due to confined boiling has also been reported to occur in ablation using continuous irradiation at relatively low irradiances [63,64].

In the above scenario, little vaporization occurs prior to the onset of ablation because bubble growth is impeded by the necessity to deform the tissue matrix. For

mechanically weak tissues, much of the ejected mass consists of tissue that is fragmented and accelerated by the phase explosion. In these cases, the ablation enthalpy can thus be considerably smaller than the vaporization enthalpy of water. However, for tissues that possess a strong ECM (e. g., skin), temperatures of 400°C to 700°C are required to produce a saturation vapour pressure exceeding the ultimate tensile strength to initiate ablation [36]. Under these conditions, the ablation enthalpy often exceeds the vaporization enthalpy of water.

The tissue matrix retains its mechanical integrity during nanosecond or microsecond laser exposures even for temperature rises of several hundred degrees. There are several factors that are responsible for this. First, as discussed in Sect. 2, the temperatures required for disintegration of the matrix increase strongly as the duration of heat exposure decreases. Second, the application of tensile stresses to collagen fibrils stabilizes the helical architecture and results in a significant increase of the denaturation temperature. Thus the generation of tensile stresses resulting from pulsed laser heating is expected to further stabilize a collagen ECM with respect to possible collagen denaturation. Third, as discussed in Sect. 2, the extreme strain rates produced by pulsed laser ablation processes ($\sim 10^5$–$10^7\,\mathrm{s}^{-1}$) likely increase the UTS of the tissue matrix. Thus ablation does not involve a "liquefaction" of the tissue as assumed in earlier models [65,66] but proceeds via the ejection of tissue fragments driven by the vaporization of tissue water.

For low rates of volumetric energy deposition ($\mu_a I$) and high number densities of heterogeneous nuclei, the nature of the ablation process is largely independent of tissue mechanical properties. The laser irradiation will initially heat the tissue under equilibrium conditions at constant pressure ($1 \rightarrow 2$) and then continue on the binodal until the ultimate tensile strength of the tissue is reached, resulting in explosive material removal. However, if the tissue is heated rapidly and/or a small number density of heterogeneous nuclei are present, the tissue water will be driven into a metastable state and a phase explosion will be induced when the spinodal limit is reached ($1 \rightarrow 2 \rightarrow 3 \rightarrow 4' \rightarrow 5$). The subsequent evolution of the process now depends on the mechanical properties of the tissue. Immediate material ejection will result for tissues that are unable to withstand the stresses and deformations associated with the phase explosion. However, tissues possessing high collagen content, and thus high UTS, will not fail mechanically due to the phase explosion. The laser irradiation will then drive a confined boiling process as indicated by path ($5 \rightarrow 6$) in Fig. 10 until the tissue ruptures at higher vapour pressures resulting in material removal. For pulsed ablation of skin, surface temperatures of 400–750 °C have been measured [36], indicating that the dynamic tensile strength of the tissue matrix is higher than the pressure at the critical point ($p_c = 22.09\,\mathrm{MPa}$).

4.6 Vapour explosion and photothermal dissociation of the tissue ECM

For ablation using nanosecond pulses, the volumetric energy densities achieved in the tissue water usually exceed the vaporization enthalpy of water [37,41]. Under these conditions, the liquid water is completely transformed into vapour in a process termed "vapour explosion". Moreover, at temperatures exceeding ~ 1000°C the constituent molecules of the ECM are thermally dissociated into volatile fragments. Energetically,

these processes result in an ablation enthalpy higher than the vaporization of water and are "less efficient" than phase explosion or confined boiling because they do not involve the ejection of condensed material.

An important factor contributing to the high volumetric energy densities achieved in nanosecond laser ablation is the recoil pressure produced by the ablation of superficial layers of the target. At the beginning of laser exposure, ablation is governed by non-equilibrium surface vaporization (Sect. 4.2). During this phase the recoil pressure is relatively small and a phase explosion occurs as soon as the temperature reaches the spinodal limit (Sect. 4.4). With the onset of the phase explosion, the ablation has transformed from a surface-mediated to a volumetric process resulting in a rapid increase of the recoil stress. This large compressive stress inhibits the ablation of deeper tissue layers until the volumetric energy density is sufficiently high to cause a phase transition that can overcome these higher pressures. Therefore the phase transition of subsurface tissue layers will be more vigorous than the initial surface vaporization because a larger volumetric energy density is required to initiate the phase change process. This, in turn, produces a higher recoil pressure that impedes ablation in deeper tissue layers until even higher volumetric energy densities are reached. At any given depth, ablation starts as soon as the vapour pressure exceeds the recoil pressure resulting from the explosive removal of more superficial layers. This results in a "positive-feedback" process in which the volumetric energy density and pressure values required for the onset of ablation at deeper tissue layers will continue to increase as long as the laser irradiance is increasing within the laser pulse. After the peak irradiance of the laser pulse has passed, the volumetric energy density and pressure at the target surface will decrease while the ablation front continues to propagate into the target. The ablation process becomes most vigorous shortly after the peak of the laser pulse as that is when the volumetric energy density in the target reaches a maximum value [37]. Since the evolution of thermodynamic states within the target is determined both by the incident laser irradiation and by the recoil produced by the ablation plume, ablation will likely continue well beyond the end of the laser pulse [6,37,67]. When the volumetric energy density in the target drops below the value required for thermal dissociation of the tissue matrix, the ejection of particulate tissue fragments will commence [37,67]. Ablation ceases when the vapour pressure within the tissue falls below the ultimate tensile strength of the tissue matrix that itself is influenced by the local denaturation kinetics (Sect. 5).

4.7 Effect of stress confinement on the ablation process

When performing ablation under conditions of stress confinement, the thermoelastic stresses modify significantly the phase transition processes that drive material removal. As discussed in Sect. 3, the thermoelastic stress wave propagation results in both compressive and tensile components (Fig. 5).

The tensile thermoelastic stress waves can produce material ejection at temperatures less than 100°C in liquids with heterogeneous vapour/cavitation nuclei and in mechanically weak tissues such as liver. This phenomenon has been compared to back surface spallation resulting from high-pressure impact [68] and investigated in several studies [69,70]. Nevertheless, temperatures above 100°C are usually necessary

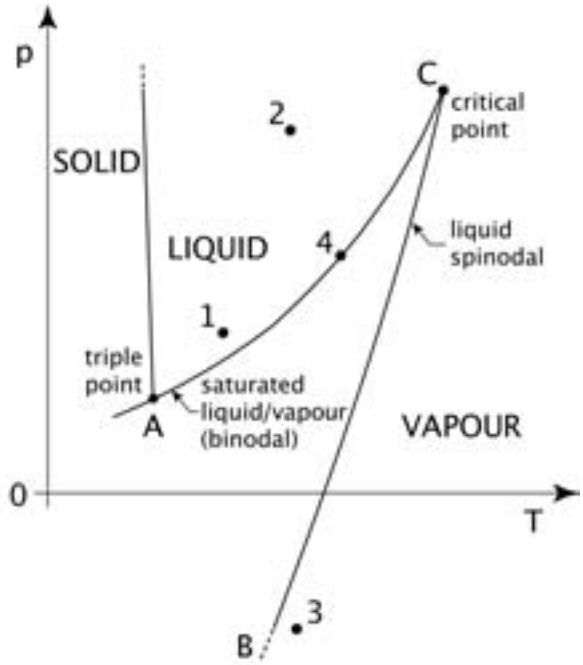


Figure 11. Path taken through the p vs. T projection of the thermodynamic phase diagram for a temperature rise above 100°C under stress confinement conditions. The transition $1 \rightarrow 2$ corresponds to the heating phase that is coupled with the generation of compressive stress. The transition $2 \rightarrow 3$ corresponds to the passage of the tensile stress wave that leads to a crossing of the spinodal limit resulting in phase separation. After the passage of the stress wave, the system reaches point 4 that corresponds to explosive boiling into the large number of bubbles produced shortly before.

to initiate efficient tissue removal [69]. The important influence of tensile stress waves at temperatures above 100°C has only recently attracted attention [6,42]. Figure 11 shows the path taken through the p–T phase diagram for a temperature rise above 100°C produced by laser irradiation with stress confinement. A heating phase ($1 \rightarrow 2$) coupled with the generation of compressive stress is followed by the passage of a tensile stress wave ($2 \rightarrow 3$) that leads to a crossing of the spinodal limit, resulting in a phase explosion. The passage of the stress wave is followed by explosive boiling into the large number of bubbles produced shortly before (point 4) resulting in vigorous material ejection. This ejection occurs at temperatures lower than the spinodal temperature at atmospheric pressure ($T = 305$°C) because, as shown in Fig. 11, the spinodal temperature is reduced with the decrease in pressure provided by the tensile thermoelastic stresses. It is important to note that the sequence of events described above may occur not only during surface ablation in a gaseous environment but also when laser pulses are focussed into a transparent material. They are the basis of the high precision in femtosecond laser nanosurgery of cells [7].

At volumetric energy densities in excess of the spinodal limit at ambient pressure, i. e., for $T > 300°C$, the superheated liquid is unstable and the onset of explosive ablation need not be initiated by the tensile component of the thermoelastic stress. Nevertheless, the thermoelastic stress transient can still contribute to material removal. The magnitude of thermoelastic transients produced by a given temperature rise under conditions of stress confinement is much larger than the saturation vapour pressure resulting from the same temperature rise and for $T > 1000°C$ may well exceed 1 GPa. The compressive component of the thermoelastic stress wave upon propagation will develop into a shock wave. The propagation of this shock wave into the depth of the target along with energy dissipation at the shock front [71,72] results in tissue heating at locations beyond those heated directly by the laser irradiation and subsequent heat diffusion. Shock wave propagation thus serves as a form of convective heat transfer that extends the ablation depth and increases ablation efficiency [73]. Experimental evidence for shock wave induced phase changes of water after laser-induced breakdown was provided by Vogel and Noack [74]. For pulsed laser surface ablation, temperatures in the shock wave region will usually be below the spinodal limit since a pressure jump in the neighbourhood of 5 GPa is required to heat water from room temperature to 300°C [71]. Nevertheless, the temperature rise can result in ablation because the tensile component of the thermoelastic stress that follows the shock wave will catalyze an explosive boiling process as described above. Convective heat transfer will become important for ablation only for sufficiently large volumetric energy densities and for very high degrees of stress confinement, i. e. mainly for ultrashort laser pulses. We conclude that regardless of the volumetric energy density, stress confinement invariably serves to lower the ablation threshold and increase ablation efficiency [6,42,69,73,75].

5 Ablation plume dynamics

The phase transitions described in the previous section drive the formation of a plume consisting of material removed from the ablation site. Usually, the ablation dynamics and plume formation is not governed by just a single type of phase transition but by an interplay of different transitions occurring at the target surface and in its bulk. Moreover, the type and strength of the phase transition may change during the laser pulse depending on the volumetric energy densities reached at each target location when the phase change occurs. The characteristics of the ablation plume reflect the underlying ablation dynamics and its analysis provides the insight necessary to draw conclusions about the phase transitions involved in a given ablation event. Furthermore, the plume dynamics influence the ablation process in various ways. The primary ejection of ablation products perpendicular to the tissue surface induces a recoil pressure that may produce additional, secondary material expulsion and cause collateral effects in the bulk tissue. Flow components parallel to the tissue surface that develop at later times may result in a redeposition of ablated material. Scattering and absorption of the incident light by the ablation plume reduce the amount of energy deposited in the target and limit the ablation efficiency at high radiant exposures.

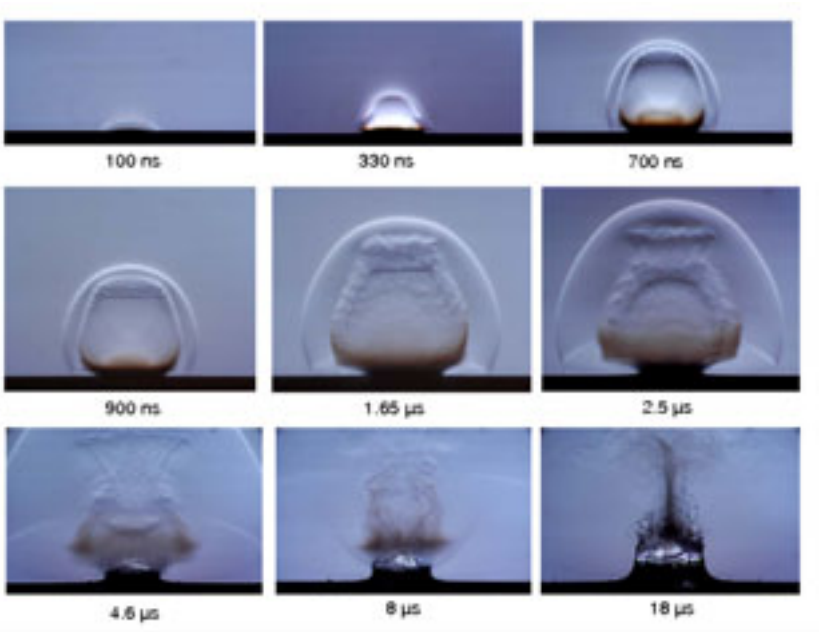


Figure 12. Early phase of water ablation by a Q-switched Er:YAG laser pulse of 70 ns duration, photographed using a novel white light Schlieren technique [76]. The irradiated spot size was 700 μm, the radiant exposure 2.8 J/cm^2 (25× ablation threshold). All times refer to the beginning of the laser pulse. The dynamics is characterized by vapour plume formation, the emission of external and internal shock waves, droplet ejection, and the onset of recoil-induced material expulsion.

To date, most investigations of the plume dynamics and acoustic phenomena associated with pulsed laser ablation of biological tissues have been performed experimentally by time-resolved photography, probe beam deflectometry, and spectroscopic techniques as reviewed in Refs. [6] and [76]. Here, we focus on the description of the plume dynamics itself rather than on the techniques of investigation. We first discuss the dynamics for water ablation and then progress to the more complicated case of tissue ablation where the primary ablation process and recoil-induced material expulsion are modified by the tissue matrix.

5.1 Primary material ejection in nanosecond ablation

For Q-switched laser pulses of 50–100 ns duration, the rate of energy deposition is extremely large. Close to threshold, the ablation process for liquids such as water is typically characterized by non-equilibrium mass transfer [52] at the target surface followed by a phase explosion of the superficial liquid layer [37]. However, when pulse energies well above the ablation threshold are used, large volumetric energy densities are produced in the target material that result in an ablation process characterized by more vigorous types of phase transitions. To illustrate this, Fig. 12 shows the sequence

of events in the early phase of Q-switched Er:YAG laser ($\lambda = 2.94\,\mu$m) ablation of water for a radiant exposure of $2.8\,\mathrm{J/cm^2}$, $\approx 25\times$ the ablation threshold. The ablation dynamics is characterized by a succession of explosive vaporization, shock wave emission, and ejection of very fine droplets. The plume remains fairly small until shortly after the peak intensity of the laser pulse, but then rapidly expands. This means that the main part of the ablated material is ejected towards the end and after the laser pulse. The layered structure of the plume reveals that different types of phase transition follow each other while the ablation front propagates into the target. The fact that the top part of the plume is completely transparent indicates that the volumetric energy density in the superficial target layers is larger than the vaporization enthalpy of water at room temperature under atmospheric pressure ($\varepsilon = 2.59\,\mathrm{kJ/cm^3}$). Therefore, this entire liquid volume is transformed into vapour in a "vapour explosion". When the ablation front has reached a depth where the energy density becomes smaller than the vaporization enthalpy of water, the superheated tissue water starts to decompose into vapour and liquid in a phase explosion, and droplet ejection commences. Droplet ejection is first visible after ≈ 700 ns and lasts for a few microseconds. The droplets cannot be resolved on the photographs and appear as a reddish haze. The reddish color indicates that the droplet size is sufficiently small to cause Rayleigh scattering by which blue light is scattered much stronger than red light [77]. As a consequence, the red spectral components of the illumination dominate the light that passes through the imaging optics. While the droplet ejection still continues, an indentation of the water surface forms and a "splash" region develops at the periphery of the ablation spot due to the recoil pressure produced by the phase transitions (see Sect. 5.3 below).

When soft tissues are ablated at moderate radiant exposures, the entire ablation plume consists of tissue fragments, as illustrated in Fig. 13(b) for Er:YAG laser ablation of liver at a radiant exposure of $1.4\,\mathrm{J/cm^2}$. At the same radiant exposure, the top layer of a water target is already completely vaporized and thus transparent as shown in Fig. 13(a). At a larger radiant exposure of $5.4\,\mathrm{J/cm^2}$ (Fig. 13(c)), the top part of the plume becomes transparent for both water and liver ablation, and particulate fragments are ejected only after about 200 ns. The sequence of gaseous ablation products followed by particulates could be visualized only by means of a photographic setup suited for detecting phase objects. In previous studies only the particulate fragments were observed and it was concluded mistakenly that the ablation process commences well after the end of the laser pulse [78]. In reality, the transparency of the top part of the plume indicates that during the initial ablation phase tissue water is completely vaporized and biomolecules are thermally dissociated into volatile fragments, which occurs at temperatures above 1000°C. For the liver target, the subsequent ejection of larger, non-transparent tissue fragments is driven by a phase explosion of the tissue water. The pressure developed during the phase separation suffices to rupture the weak tissue matrix in liver parenchyma (Sect. 4.5). The ejection ceases when the ablation front reaches a depth where the temperature drops below the stability limit of the superheated tissue water. The different optical appearance of the transparent and opaque parts of the ablation plume is due to differences in molecular composition and particle size distribution but not necessarily indicative for disparities in the average mass density.

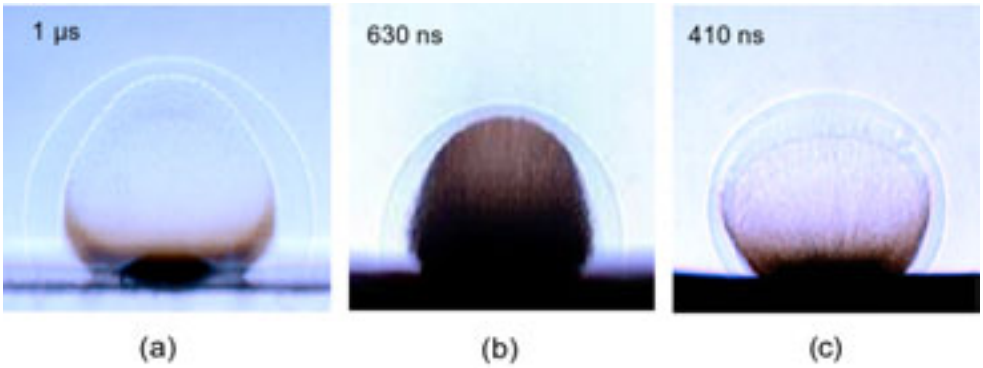


Figure 13. Q-switched Er:YAG laser ablation of (a) water at $\Phi = 1.4\,\mathrm{J/cm^2}$, (b) liver at $\Phi = 1.4\,\mathrm{J/cm^2}$, and (c) liver at $\Phi = 5.4\,\mathrm{J/cm^2}$. The plume consists of water vapour (top) and a droplet/vapour mixture in (a), tissue fragments in (b), and dissociated biomolecules (top) and tissue fragments (bottom) in (c). The volumetric energy densities averaged over the optical penetration depth are $\approx 5.2\,\mathrm{kJ/cm^3}$ in (a), $\approx 4\,\mathrm{kJ/cm^3}$ in (b), and $\approx 9\,\mathrm{kJ/cm^3}$ in (c).

For the ablation of skin at large radiant exposures, a similar sequence of biomolecule dissociation followed by ejection of tissue fragments was observed [37]. However, in this case the ejection of tissue fragments occurred over a shorter time interval than for liver. Ablation ceased when the ablation front reached a depth where the vapour pressure dropped below the tensile strength of the extracellular tissue matrix. Nevertheless, fragment ejection was found to continue for several microseconds after the laser pulse while the tissue matrix is increasingly weakened by thermal denaturation. Generally, the size of the ejected tissue particles is small at early times after the laser pulse and increases with time [37,78]. The entire sequence of phase transitions occurring during water and tissue ablation is summarized in Fig. 14.

Since ablation becomes a volumetric process as soon as the spinodal limit is exceeded and a phase explosion sets in (Sect. 4.4), it is not self-evident why large volumetric energy densities sufficient for a vapour explosion and dissociation of biomolecules should be reached in pulsed laser tissue ablation. However, one needs to consider that the recoil stress produced by the phase transitions of the uppermost tissue layers delays the phase transitions in underlying layers because the spinodal temperature increases with increasing pressure (see Fig. 6). The ongoing absorption of laser energy into the underlying layers can thus drive the thermodynamic state into the supercritical regime. Even larger recoil stresses are produced when these layers are ablated, and the phase transitions in deeper layers are delayed even more. This "positive-feedback" process continues at least until the intensity peak of the laser pulse is reached after which a relaxation process resulting in explosive ablation commences and continues for several microseconds after the end of the laser pulse. The energy densities generated during the runaway process are in the order of $10\,\mathrm{kJ\,cm^{-3}}$ [37] and give rise to recoil pressures of several hundred MPa (Sect. 4.3).

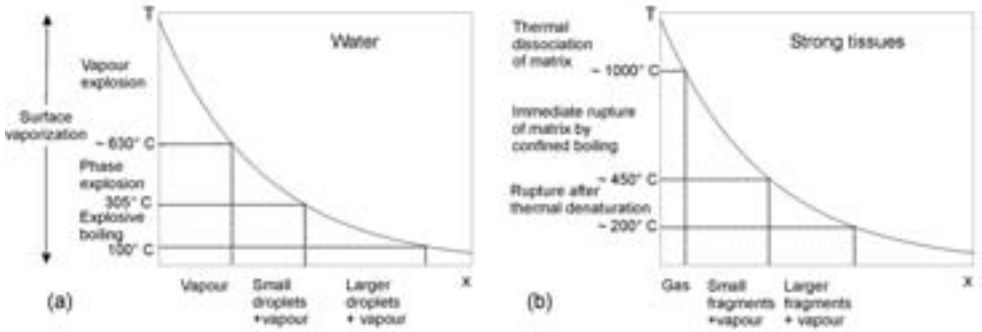


Figure 14. Sequence of phase transitions and corresponding plume constituents in ablation at radiant exposures well above threshold for (a) water ablation, and (b) ablation of mechanically strong tissues such as, for example, skin.

The high volumetric energy density in the target material produced in Q-switched Er:YAG laser ablation results in a very large initial expansion velocity of the ablation plume that drives the emission of an equally fast shock wave. Shock front velocities are usually on the order of 2000–4000 m/s for both IR and UV wavelengths [37, 79–81], i. e., they reach values up to Mach 12. Measured shock wave and plume velocities correlate with the water content of the samples because lower water content results in smaller volumetric energy densities and less vigorous ablation. By contrast, the velocity of particulate fragments is larger for mechanically strong tissues (up to 1700 m/s for skin) than the velocity of droplets ejected in water ablation (up to 280 m/s) [37]. This is because the temperature required for thermal dissociation of the tissue matrix into volatile products is higher than the temperature required for complete vaporization of water. Therefore, tissue fragments become visible early in the ablation process when the ablation front has reached a depth at which the temperature is below the level required for thermolysis. At this time, the pressure driving the ejection is still very high. By contrast, droplet ejection starts only once the temperature at the ablation front has reached a lower level corresponding to the onset of a phase explosion. This results in smaller velocities for the droplet ejection.

The ablation plume exhibits complex dynamics. The plume expansion is nearly spherical during the initial phases of expansion but begins to propagate preferentially in the forward direction after 1–2 μs. For small radiant exposures, the interaction of the piston-like forward movement with the ambient air at rest results in ring vortex formation [37,82]. For larger radiant exposures, a region of high density and pressure is created at the contact front between plume and surrounding air. The molecules and molecular clusters propagating with the plume possess a non-zero average velocity. When they collide with air molecules that are, on average, at rest, they are partially reflected back into the plume. As visible in Fig. 12, this reflection leads to the formation of an internal shock wave that begins to propagate toward the target surface when the rarefaction from the plume expansion has reduced the

pressure in the plume considerably below its initial value [37,83,84]. The internal shock interacts with the particles and droplets of the plume and deforms the shape of the particle cloud during a time interval lasting about 10 μs. Due to the heating at the shock front, the passage of the internal shock wave through the reddish droplet cloud results in their vaporization.

The propagation of the shock front after a strong explosion in a homogeneous atmosphere was first theoretically described by Taylor [85] and Sedov [86,87] and, using a higher-order approximation, by Sakurai [88,89]. These theories neglect the mass of the gas and debris driving the shock wave and are thus valid only once the shock wave has swept over a mass of atmospheric gas much greater than the mass in which the energy was initially concentrated. Various authors have later obtained solutions for the mass-dependent flow regime [90–92] and simple analytic solutions are available for some limiting cases. When the mass of the gas encompassed by the shock wave is much greater than the initial ablated mass and the pressure driving the shock is much greater than the atmospheric pressure ahead of the shock front, the position $R(t)$ of a spherical shock wave is governed by [93]

$$R(t) = \xi (E_0/\rho_0)^{1/5}\, t^{2/5}\,, \tag{9}$$

and that of a planar shock wave such as emitted from a large irradiated spot size by

$$R(t) = \xi (E_0/\rho_0)^{1/3}\, t^{2/3}\,. \tag{10}$$

Here E_0 is the energy driving the explosion, ρ_0 the density of the undisturbed gas, and ξ is a constant that is a function of the specific heat capacity ratio γ of the gas. The peak pressure produced scales proportional to $E_0^{2/5}$. Once the shock wave pressure becomes comparable to the ambient pressure, its propagation is better described by the Jones approximation [94,95]. When the mass of the material removed is very large or the background pressure very low (including vacuum), the motion of a planar shock wave can be described by [96]

$$R(t) = \xi (E_0/\rho_0)^{1/2}\, t\,. \tag{11}$$

A comparison of experimental $R(t)$ data with Eqs. (9–11) allows an assessment of the transduction of laser pulse energy into blast wave energy E_0 [95,97].

More refined numerical simulations by Brode [84] and the analytical treatment by Arnold and co-workers [83] include the spherical movements of the external shock front, the contact front between plume and ambient gas, and the internal shock front within the plume. Recently, Chen and co-workers [98] presented a model for the propagation of the external shock wave propagation in atmospheric pressure laser ablation of water-rich targets that incorporates the nonlinear absorption of water and the phase explosion due to superheating. The model predicts a succession of an initially slow plume emission followed by a vigorously accelerated expansion, in good agreement with the experimental results of Apitz and Vogel [37] and with the views presented above.

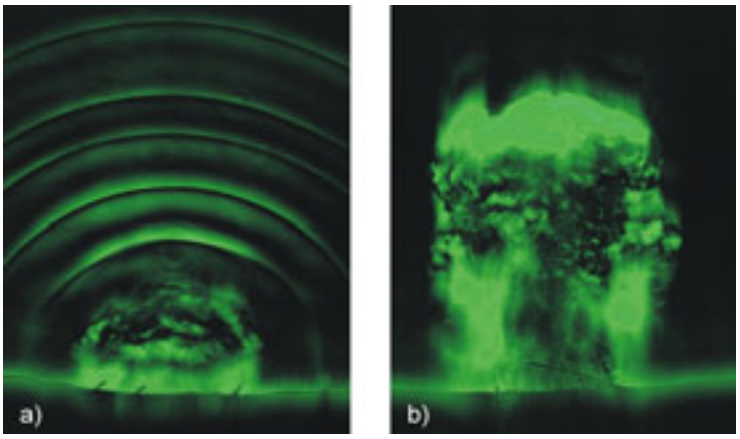


Figure 15. Dark-field Schlieren images of the acoustic transients and ablation plume during skin ablation with a 200-μs Er:YAG laser pulse ($\Phi = 20\,\mathrm{J/cm^2}$, spot size 2.3 mm) photographed (a) 22.4 μs and (b) 40 μs after the onset of the laser pulse. The images show acoustic transients arising from individual spikes in the free-running laser irradiation, and the plume containing vapour and tissue fragments. (Reprinted from Ref. [76] with permission. Copyright 2006 Optical Society of America).

5.2 Primary material ejection in microsecond ablation

Free-running lasers typically provide pulse durations longer than 100 μs. Thus, unlike nanosecond ablation, plume formation and expansion occurs largely during the laser irradiation. As a result, the ablation plume influences the energy deposition of the laser radiation into the tissue target, and the plume dynamics is also influenced by the interaction of the laser beam with the ejected material. Nevertheless, the succession of a sub-ablative phase, development of a vapour plume, and material ejection is similar as with nanosecond pulses even though it occurs on a much longer time scale [99]. However, the heating rates available from microsecond laser pulses are generally much smaller than those available from nanosecond laser pulses of moderate to high radiant exposures. These lower heating rates are not sufficient to generate the temperatures necessary to dissociate ECM molecules and are only able to produce supercritical water at very large radiant exposures.

Free-running laser emission is characterized by intensity fluctuations during the laser pulse ("spiking" behaviour). These intensity peaks modulate the vaporization and material ejection rates [78,100] as well as the emission of acoustic transients generated during the ablation process [101,102]. The intensity spikes of the laser pulse are coupled with the generation of individual transients as shown in Fig. 15.

The mechanisms leading to material ejection are the same as for nanosecond pulses: a phase explosion for mechanically weak materials and a succession of phase explosion and confined boiling for mechanically stronger tissues. Previously it was believed that the generation of a phase explosion requires pulse durations in the nanosecond range [41]. However, using time-resolved photography, Nahen and Vogel [99] demon-

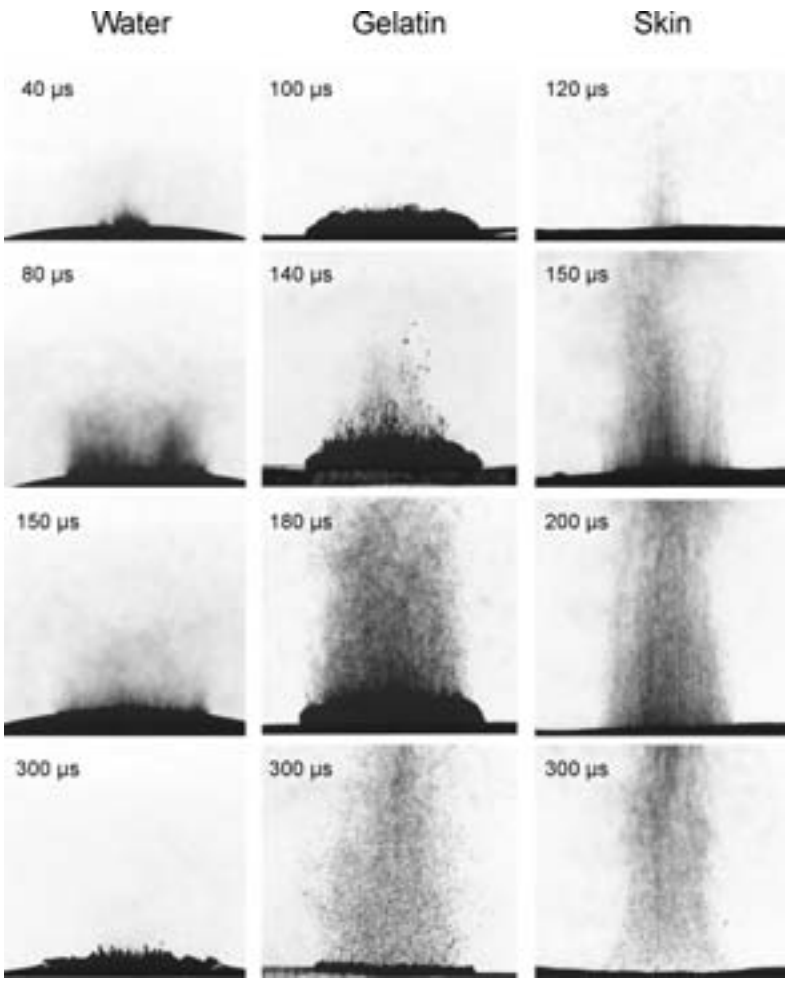


Figure 16. Dynamics of Er:YAG laser ablation of water, gelatin with 70% water content, and skin using a radiant exposure of $4.6\,\mathrm{J/cm^2}$, 5 mm spot size, and $200\,\mu$s pulse duration. The times after the beginning of the laser pulse when the photographs were taken are indicated on the individual frames. Note the increasing delay in the ejection of particulate matter with increasing mechanical strength of the target.

strated that a phase explosion can also be produced with pulse durations on the order of $200\,\mu$s. This is shown in Fig. 16 that compares the ablation dynamics for Er:YAG laser irradiation of water, gelatin and skin using identical radiant exposures. The rapid droplet ejection during Er:YAG laser ablation of water can only be produced by a phase explosion because in the absence of stress confinement no other mechanism gives rise to a material ejection perpendicular to the water surface. In gelatin, a phase explosion occurs at the same time as in water. However, the phase explosion only deforms the gelatin surface without rupturing it, and fracture of the

gelatin surface and rapid particle ejection are observed only after a further pressure build-up through confined boiling (Sect. 4.5). The material ejection during skin ablation is also characterized by a phase explosion followed by confined boiling. However, the higher mechanical strength of skin causes a further delay of material ejection compared to gelatin. It is important to note that both for skin and gelatin targets fragments are ejected in the form of solid particles. The absence of droplet-like ejecta indicates that gelatin exposed to temperatures near the spinodal limit does not melt within 200 μs, even though it melts at 60°C for sufficiently long heat exposures. This finding is consistent with the strong increase in denaturation temperature for very short exposures that was discussed in Sect. 2.

Initial material ejection velocities observed for microsecond laser ablation are roughly one order of magnitude lower than those reported for nanosecond ablation [103]. For free-running pulses, an increase of the radiant exposure results in an earlier onset of the material ejection but does not change the ejection velocity significantly. It is only for very large radiant exposures in which the first intensity spike of the free-running pulse provides a dose in excess of the ablation threshold that an increase of the ejection velocity is observed. By contrast, for nanosecond exposures an increase of the radiant exposure is always coupled with an increase of the volumetric energy density that translates directly into a higher temperature, pressure, and ejection velocity.

In both Q-switched and free-running laser ablation of soft tissues, material ejection continues for a considerable time following laser irradiation that can last up to several milliseconds [78–80,82,99,104]. In general, post-pulse ablation lasts longer for mechanically weaker tissues, larger radiant exposures, and larger laser beam diameters. One possible driving force for the continuation of the ablation process after the end of the laser pulse is the heat retained in the tissue. A progressive weakening of the tissue matrix through thermal denaturation enables a propagation of the ablation front until the vapour pressure in the residual tissue drops below the UTS of the weakened tissue matrix. Another very important source of post-pulse ablation are hydrodynamic phenomena such as recoil stress-induced material expulsion.

5.3 Recoil stress and secondary material ejection

Both the rapidly expanding vapour plume and the ejected particles generate recoil stresses that impart momentum to the tissue. The linear momentum per unit area of the ablated material l is the time integral of the recoil stress σ_{rec} at the target surface

$$l = \int_0^\infty \sigma_{\mathrm{rec}}(t)\,\mathrm{d}t\,. \tag{12}$$

A derivation of the peak recoil stress requires assumptions on the nature and duration of the ablation process. Various authors have presented solutions for the peak stress amplitude produced by a continuous vaporization process [40,66,105,106], and by explosive ablation where the entire laser pulse is deposited prior to the onset of material removal [41,107].

Experimental values for the recoil stress produced by nanosecond laser ablation have been obtained through direct pressure measurements using piezoelectric trans-

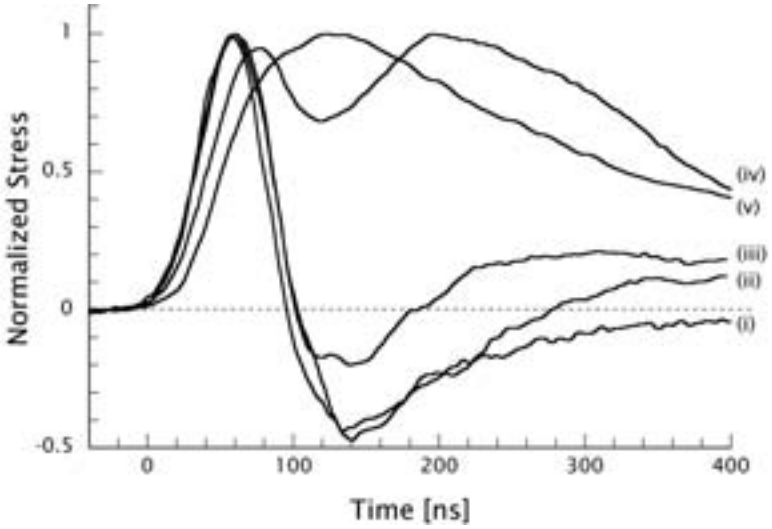


Figure 17. Stress transients resulting from TEA CO_2 laser (t_p = 30 ns) irradiation of porcine dermis for radiant exposures below the ablation threshold (i), at threshold (ii), and above threshold (iii)–(v). Radiant exposures below threshold produce bipolar thermoelastic stress transients. For radiant exposures equal to and above threshold, a compressive pulse is produced by the ablative recoil. (Adapted from Ref. [41] with permission. Copyright 1996 Biophysical Society)

ducers [40,41,108,109], and, for water ablation, through analysis of the speed of the recoil-induced shock wave [37]. Peak pressures range from a few MPa at the ablation threshold up to several hundred MPa for radiant exposures well above threshold. For free-running microsecond laser pulses, average stress values during the laser irradiation have been determined through measurement of the recoil momentum using the ballistic pendulum method [65,106], while the peak stress amplitudes produced by the intensity maxima of the free-running pulses were obtained via transducer measurements [109]. Peak values of recoil stress produced during cornea ablation using free-running Er:YSSG laser irradiation at a radiant exposure of 50 J/cm^2 amounted to 2 MPa [109] while the average pressure value for skin ablation at the same radiant exposure was only 0.3 MPa [106].

In stress-confined tissue ablation, the compressive recoil stress transient is superimposed on a bipolar thermoelastic transient [110]. Figure 17 demonstrates the transition from a bipolar stress transient for radiant exposures below the ablation threshold to a monopolar compressive transient when the ablation threshold is exceeded. This transition and the corresponding increase in peak pressure is a sensitive method for the determination of the ablation threshold [40,41,108].

The recoil stress produced by both vaporization and material ejection in the primary ablation phase can induce a secondary material expulsion process that leads to a strong increase of the ablation efficiency [37,105]. Recoil-induced material expulsion is most pronounced during ablation of liquids and mechanically weak tissues.

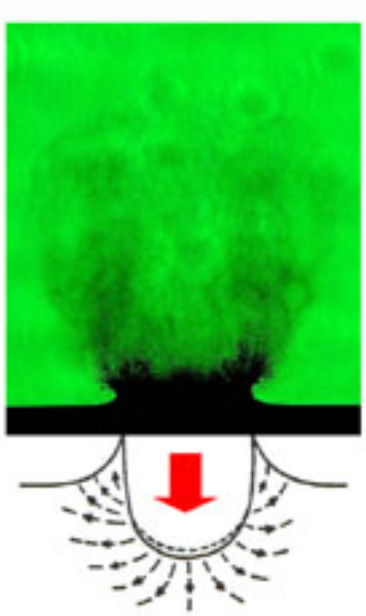


Figure 18. Recoil-induced material expulsion in water ablation by 200-μs Er:YAG laser pulses, together with a schematic illustration showing the pathlines of the ejected material fragments. The lateral component of the recoil-induced flow collides with the surrounding fluid that is at rest, thus producing an upward directed splash.

Material will be ejected whenever the recoil stress component in the radial direction exceeds the mechanical strength of the tissue, as illustrated in Fig. 18.

The sequence of primary material ejection and recoil-induced material expulsion is shown in Fig. 19 for free-running and Q-switched Er:YAG laser ablation of liver. While the primary material ejection visible at short delay times takes place across the entire ablation area, recoil-induced expulsion occurs preferentially at the ablation crater rim and includes the ejection of tissue fragments much larger than those ejected during the initial phase explosion. The recoil-induced ejection dynamics resembles the surface indentation and subsequent "splash" produced by the impact of liquid droplets on bulk liquids that has already been investigated in considerable detail [111,112]. The mass expelled at later times far exceeds the mass ejected during the primary ablation phase. However, the velocity of the ejecta is considerably slower.

Recoil-induced material expulsion begins after the primary ejection process, requires a radiant exposure well above the ablation threshold, and provides an increase of the ablation efficiency. A marked increase of the ablation efficiency at a certain radiant exposure has been observed for weak tissues as liver and myocardium as well as for gelatin with high water content but not for tissues with greater mechanical strength such as skin [37,113]. Remarkably, no recoil-induced ejection was observed in skin ablation using Q-switched Er:YAG laser pulses even when the recoil stress was about 50 times larger than the quasi-static ultimate tensile strength of skin [37].

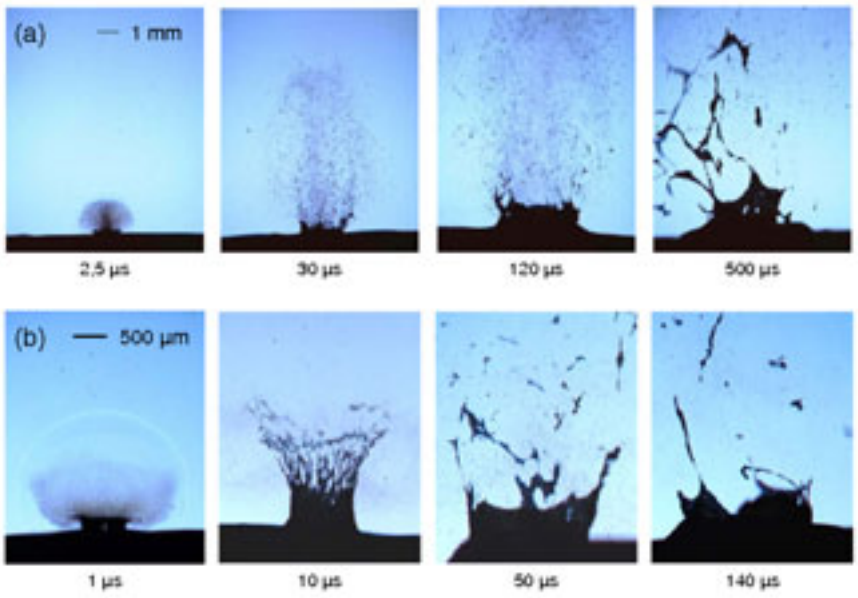


Figure 19. Recoil-induced material expulsion for liver ablation by (a) 200-μs Er:YAG laser pulses at 100 J/cm^2 radiant exposure and 1.1 mm spot size, (b) 70-ns Er:YAG laser pulses at 5.4 J/cm^2 radiant exposure and 0.5 mm spot size. The primary material ejection produced by the phase changes in the target is also visible in all images of (a) and in the first image of (b).

To understand this discrepancy, we must first consider that the recoil-induced tensile and shear stresses that contribute to tissue fracture may be considerably smaller than the measured compressive recoil stress. Moreover, as discussed in Sect. 2, the dynamic tensile strength of tissue at the extreme strain rates produced in pulsed laser ablation is much higher than the quasi-static values for the UTS found in the literature. Finally, tissue fracture will only occur at sufficiently large strain that may not be achieved by stress transients of very short duration [6,22].

For mechanically weak tissues, the recoil-stress-induced material expulsion produces craters with a depth much larger then the optical penetration depth and a diameter much larger than the irradiated spot size, as shown in Fig. 20(a,b). For mechanically strong tissues, the recoil stress does not lead to material expulsion. However, it can produce tissue tearing at the sides of the ablation crater as seen in Fig. 20(c). The cracks and tearing patterns arise preferentially along morphological structures with reduced mechanical strength such as the transitions between corneal lamellae, sinusoid spaces holding blood between plates of cells in liver tissue, and their orientation is also influenced by the weakness of the longitudinal strength of blood vessels compared to their circumferential strength [106]. Tissue tearing at the rim of ablation craters was not observed for skin due to its three-dimensional collagen network that results in an approximately isotropic UTS.

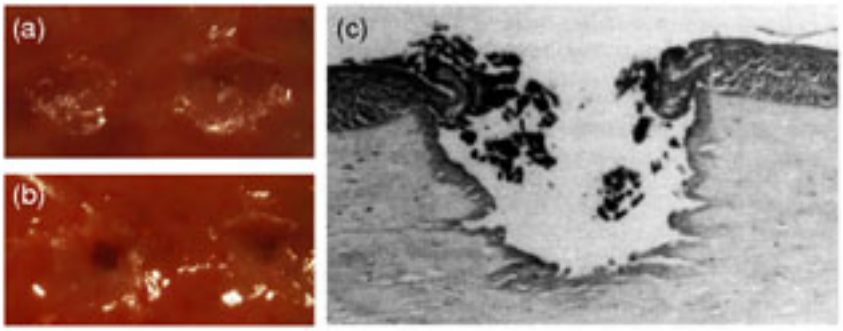


Figure 20. (a,b) Craters produced during liver ablation by 200-μs Er:YAG laser pulses of 32 J/cm^2 and 50 J/cm^2 radiant exposure (2.5 and 1.6 mm spot size, respectively). (c) Histologic slide showing an ablation crater in bovine cornea produced by an Erbium:YSGG laser pulse with $t_p = 250\,\mu$s and $\Phi_0 = 100$ J/cm^2. The tissue around the ablation crater exhibits a 25–50 μm zone of thermal damage (dark) and mechanical tearing between the corneal lamellae. (c) is reprinted from Ref. [106] with permission. Copyright 1993 Optical Society of America).

Recoil-induced stress transients can produce also more subtle forms of collateral tissue damage further away from the irradiation site. Putative photoacoustic damage created during ArF-excimer ($\lambda = 193$ nm) laser ablation of skin was described by Watanabe and co-workers [114] and Yashima and co-workers [115]. The formation of tensile stress with an amplitude of 3.5 MPa inside the eyeball through diffraction of the recoil stress wave produced during ArF-excimer laser ablation of the cornea was shown by Pini and co-workers [116]. Könz and co-workers [109] demonstrated recoil-induced damage of the corneal endothelium after mid-IR laser ablation of the corneal stroma that was due to the tensile stress generated upon partial reflection of the compressive recoil stress transient at the cornea–aqueous interface.

Thus, to achieve precise and gentle tissue ablation it is not sufficient to simply select a laser wavelength with small optical penetration depth and a pulse duration providing thermal confinement. In addition, one must avoid the production of extensive recoil stresses that may degrade the quality of the ablated surface and/or induce collateral mechanical damage. This restriction imposes an upper limit for the incident radiant exposure.

5.4 Shielding and flow-induced material redeposition

Absorption, scattering, and diffuse reflection of incident laser light by the ablation plume leads to a reduction of the energy delivered to the target tissue and a reduction of the ablation efficiency. Direct measurements of the diffuse reflectance of the plume [117] and of the entire reduction of optical transmission through the plume [99] yielded values of the extinction coefficient within the plume produced by soft tissue ablation using Er:YAG laser irradiation ($\lambda = 2.94\,\mu$m) on the order of 1 cm^{-1} [99].

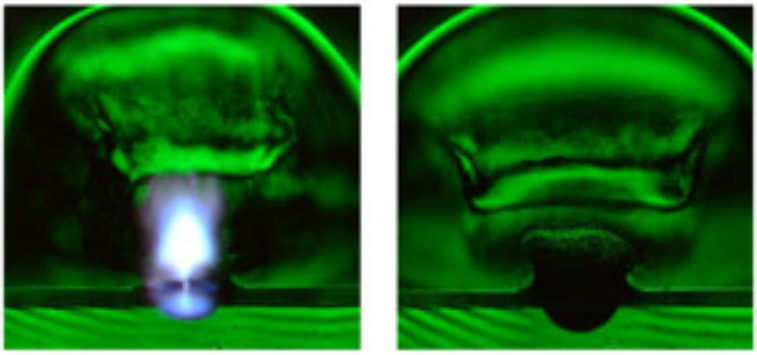


Figure 21. Dark field images of water ablation at 5.4 J/cm^2 radiant exposure with and without plasma formation (0.5 mm spot size, image taken 2 μs after the laser exposure). The plasma originates in hot spots at the water surface and grows into the incoming laser beam. The recoil-induced cavity is considerably smaller in the case with plasma formation due to the "shielding" of the target by light absorption in the plasma. (Reprinted from Ref. [37] with permission. Copyright 2005 Springer Verlag).

Plume reflectance measurements by Nishioka and Domankevitz [117] showed that shielding is strongly enhanced when a series of pulses is applied instead of single pulses. For Er:YAG laser ablation of skin at a spot size of 2 mm, Kaufmann and Hibst observed a decrease of the etch depth per pulse from 40 μm to 10 μm when the pulse repetition rate was increased from 1 Hz to 10 Hz [118]. The reduction of the ablation efficiency was attributed to increased shielding by the ablation plume. When a considerably smaller spot size is used, the lateral spread of the plume removes a larger fraction of the ablation products out of the beam path, and the etch depth does not decrease with increasing repetition rate [119].

At very high radiant exposures, plasma formation in front of the target may lead to a further decrease of the optical transmission to the target [37,40,120], as illustrated in Fig. 21. Plasma formation starts at the target surface but the plasma grows rapidly after ignition into the space in front of the target surface causing an effective shielding effect.

When radiant exposures close to the ablation threshold are used, the plume acquires a mushroom-like shape that exhibits a ring vortex at its top, a thin stem with a diameter smaller than the ablation spot and a radial flow component parallel to the surface at the foot of the plume (see also Fig. 12) [37,82,104]. The radial flow parallel to the tissue surface can result in a redistribution of ablation products across the ablation spot. For example, when performing corneal refractive surgery using large laser spot sizes, ablation rates were found to be smaller in the centre of the ablation zone than in its periphery even though the irradiance was spatially homogeneous. As a result "central islands" remained that distorted the intended refractive correction [121]. Photographic investigations of the plume dynamics revealed that the "central islands" are the result of a redeposition of ablated material after the end of the laser pulse that preferentially occurs near the stagnation point of the flow at the

centre of the ablated area [82]. A second factor contributing to the non-uniformity of the ablation rate is the attenuation of subsequent laser pulses in the centre region of the ablated area by remnants of the plume from previous pulses that preferentially stay in the vicinity of the stagnation point [82].

6 Conclusions

6.1 Ablation models

In the previous sections we established that pulsed laser ablation always consists of a sequence of different phase transition processes that occur during and after the laser irradiation. Moreover, we have shown that the initial primary material ejection is often followed by a secondary, recoil-induced, ejection. The type and vigour of the phase transition and ejection processes during an individual ablation event depend on both the laser irradiance and radiant exposure as well as on the optical and mechanical tissue properties. Thus, it is impossible to formulate a simple comprehensive ablation model that describes these different aspects of the ablation process. Nevertheless, it is useful to formulate simplified models that elucidate basic features of the ablation behaviour and parameter dependencies for specific ablation regimes.

One approach is to use metrics of the ablation process such as the threshold, enthalpy, and efficiency of ablation to predict ablation rates without reference to mechanistic aspects of the ablation process. Such heuristic models are particularly valuable to illustrate the ablation behaviour in extreme cases such as the "steady state" model for long laser pulses and the "blow-off" model for very short pulses [6]. Unfortunately, their relative success has, for a long time, obscured the real complexity of the ablation behaviour.

Analytical models that link the ablation outcome to underlying mechanisms provide more insight into the collateral damage arising from ablation than the heuristic models, but all models presented to date are applicable only over a very limited range of radiant exposures and material properties [6]. One will have to resort to computational approaches to model the full complexity of pulsed laser tissue ablation. However, much information on dynamic material properties required for a faithful modelling is still missing. These data include dynamic optical properties (absorption and scattering coefficients), thermal properties (heat capacity, Grüneisen coefficient), and mechanical properties (elastic and shear modulus, ultimate tensile strength) that depend on the magnitude and kinetics of the temperature, pressure, and strain rates achieved during the ablation process. This lack of data is now recognized to be even more important than was thought just a few years ago as recent experimental studies demonstrate that the temperatures and pressures involved in most ablation processes are more extreme than assumed previously. We now know that pulsed laser tissue ablation may be associated with a temperature rise of hundreds to thousands of K, recoil pressures of several hundred MPa, and strain rates on the order of 10^5 to $10^7\,s^{-1}$ [6,37]. Several of the above mentioned dynamic material properties are presently known only at the lower margin of this parameter space.

Molecular dynamics simulations [122–124] have yielded a wealth of information regarding the inception of phase transitions and the time-evolution of the size and

velocity distribution of the ablation products that are difficult to obtain by other means. However, enormous computational facilities, that are currently unavailable, are required to perform simulations that span the entire spatial and temporal scales usually encountered in realistic applications of tissue ablation.

6.2 Control of ablated mass and thermal and mechanical side effects

The heuristic ablation models mentioned above predict that, if the *removal of large amounts of material* is desired, the use of long laser pulses that achieve a steady-state-like ablation process will be more suitable than ablation based on a blow-off process. This arises because the ablated mass scales linearly with radiant exposure in a steady state ablation process but logarithmically in a blow-off process [6]. However, the difference between the two types of ablation process becomes less pronounced if optical shielding by the ablation products (Sect. 5.4) is significant. A steady state process is most advantageous under conditions where the absorption of the incident laser beam by the ablation plume is markedly smaller than the absorption in the target tissue present in the blow-off situation.

The most direct strategy to *control thermal side effects* involves the selection of a pulse duration that is sufficiently short to minimize heat diffusion during the laser pulse from the volume of energy deposition into the non-ablated tissue (Eq. (4)). However, similar results may also be obtained using longer pulses if the velocity of the ablation front during the laser pulse is comparable or faster than the heat diffusion into the residual tissue. A theoretical analysis of this strategy was presented by Venugopalan and co-workers [53], and experimental evidence for its validity was presented by various authors [53,118,125].

Thermal side effects can be diminished further by selecting laser pulse durations sufficiently short to provide both thermal and stress confinement. Stress confinement serves to lower the ablation threshold and increase the ablation efficiency (Sect. 4.7), and the lowering of the ablation enthalpy in the stress confinement regime reduces the residual heat in the tissue [126,127].

Following the above, high precision ablation can be achieved by selecting a laser wavelength featuring a very small optical penetration depth combined with a short pulse duration sufficient to provide thermal confinement, and, if possible, also stress confinement. However, we showed in Sect. 5.3 that one also needs to avoid the production of extensive recoil stresses to minimize *mechanical side effects*. Extensive recoil stresses may degrade the smoothness of the ablated surface and/or induce collateral mechanical damage; especially in friable tissues [37]. This restriction imposes an upper limit for the incident radiant exposure and implies that the finest tissue effects can be achieved by working close to the ablation threshold.

References

[1] L. R. Solon, R. Aronson, and G. Gould, 'Physiological implications of laser beams', Science **134**, 1506 (1961).

[2] R. F. Steinert and C. A. Puliafito, *The Nd:YAG Laser in Ophthalmology. Principles and Clinical Practice of Photodisruption* (Saunders, Philadelphia, 1986).

[3] J. M. Krauss, C. A. Puliafito, and R. F. Steinert, 'Laser interactions with the cornea', Surv. Ophthalmol. **31**, 37 (1986).

[4] A. J. Welch and M. J. C. van Gemert, eds., *Optical-Thermal Response of Laser-Irradiated Tissue* (Plenum Press, New York, 1995).

[5] M. H. Niemz, *Laser-Tissue Interactions. Fundamentals and Applications* (Springer, Berlin, 2002), 2nd ed.

[6] A. Vogel and V. Venugopalan, 'Mechanisms of pulsed laser ablation of biological tissues', Chem. Rev. **103**, 577 (2003).

[7] A. Vogel, J. Noack, G. Hüttman, and G. Paltauf, 'Mechanisms of femtosecond laser nanosurgery of cells and tissues', Appl. Phys. B **81**, 1015 (2005).

[8] R. A. J. Eady, I. M. Leigh, and F. M. Pope, 'Anatomy and Organization of Human Skin', in *Rook/Wilkinson/Ebling Textbook of Dermatology*, edited by R. H. Champion, J. L. Burton, D. A. Burns, and S. M. Breathnach (Blackwell Science, Oxford, UK, 1998), 6th ed.

[9] F. H. Silver, *Biological materials: Structure, Mechanical Properties, and Modeling of Soft Tissue* (New York University Press, New York and London, 1987).

[10] S. Nomura, A. Hiltner, J. B. Lando, and E. Baer, 'Interaction of water with native collagen', Biopolymers **16**, 231 (1977).

[11] W. F. Cheong, 'Summary of optical properties', in *Optical-Thermal Response of Laser-Irradiated Tissue*, edited by A. J. Welch and M. J. C. van Gemert (Plenum Press, New York, 1995), pp. 275–304.

[12] R. M. P. Doornbos, R. Lang, M. C. Aalders, F. W. Cross, and H. J. C. M. Sterenborg, 'The determination of in vivo human tissue optical properties and absolute chromophore concentrations using spatially resolved steady-state diffuse reflectance spectroscopy', Phys. Med. Biol. **44**, 967 (1999).

[13] T. L. Troy and S. N. Thennadil, 'Optical properties of human skin in the near infared wavelength range of 1000 to 2200 nm', J. Biomed. Opt. **6**, 167 (2001).

[14] S. Jacques, 'Role of tissue optics and pulse duration on tissue effects during high-power laser irradiation', Appl. Opt. **32**, 2447 (1993).

[15] S. A. Carp, S. A. Prahl, and V. Venugopalan, 'Radiative transport in the delta-P_1 approximation: accuracy of fluence rate and optical penetration depth predictions in turbid semi-infinite media', J. Biomed. Optics **9**, 632 (2004).

[16] R. Shori, A. A. Walston, O. M. Stafsudd, D. Fried, and J. T. Walsh, 'Quantification and modeling of the dynamic changes in the absorption coefficient of water at 2.94 μm', IEEE J. Sel. Topics Quantum Electron. **7**, 959 (2002).

[17] J. P. Cummings and J. T. Walsh, 'Erbium laser ablation: The effect of dynamic optical properties', Appl. Phys. Lett. **62**, 1988 (1993).

[18] K. L. Vodopyanov, 'Saturation studies of H_2O and HDO near 3400 cm^{-1} using intense picosecond laser pulses', J. Chem. Phys. **94**, 5389 (1991).

[19] J. T. Walsh and J. P. Cummings, 'Effect of the dynamic optical properties of water on midinfrared laser ablation', Lasers Surg. Med. **15**, 295 (1994).

[20] B. I. Lange, T. Brendel, and G. Hüttmann, 'Temperature dependence of the absorption in water of light at holmium and thulium laser wavelengths', Appl. Opt. **41**, 5797 (2002).

[21] P. T. Staveteig and J. T. Walsh, 'Dynamic 193-nm optical properties of water', Appl. Opt. **35**, 3392 (1996).
[22] F. A. Duck, *Physical Properties of Tissue* (Academic Press, London, 1990).
[23] H. G. Vogel, 'Influence of age, treatment with corticosteroids and strain rate on mechanical properties of rat skin', Biochim. Biophys. Acta **286**, 79 (1972).
[24] R. C. Haut, 'The effects of orientation and location on the strength of dorsal rat skin in high and low speed tensile failure experiments', Trans. ASME Biomed. Eng. **111**, 136 (1989).
[25] G. W. Dombi, R. C. Haut, and W. G. Sullivan, 'Correlation of high-speed tensile strength with collagen content in control and lathyritic rat skin', J. Surg. Res. **54**, 21 (1993).
[26] L. Stryer, *Biochemistry* (WH Freeman and Company, San Francisco, 1987).
[27] S. Thomsen, 'Pathological analysis of photothermal and photomechanical effects of laser-tissue interactions', Photochem. Photobiol. **53**, 825 (1991).
[28] M. E. Nimni and R. D. Harkness, 'Molecular structure and functions of collagen', in *Collagen. Vol. I Biochemistry*, edited by M. Nimni (CRC Press, Boca Raton, 1988), pp. 1–77.
[29] J. C. Allain, M. Le Lous, L. Cohen-Solal, S. Bazin, and P. Maroteaux, 'Isometric tensions developed during the hydrothermal swelling of rat skin', Connect. Tissue Res. **7**, 127 (1980).
[30] J. Kampmeier, B. Radt, R. Birngruber, and R. Brinkmann, 'Thermal and biomechanical parameters of porcine cornea', Cornea **19**, 355 (2000).
[31] M. N. Asiyo-Vogel, R. Brinkmann, H. Notbohm, R. Eggers, H. Lubatschowski, H. Laqua, and A. Vogel, 'Histologic analysis of thermal effects of laser thermokeratoplasty and corneal ablation using Sirius-red polarization microscopy', J. Cataract Refract. Surg. **23**, 515 (1997).
[32] M. Le Lous, F. Flandin, D. Herbage, and J. C. Allain, 'Influence of collagen denaturation on the chemorheological properties of skin, assessed by differential scanning calorimentry and hydrothermal isometric tension measurement', Biochim. Biophys. Acta **717**, 295 (1982).
[33] F. C. Henriques, 'Studies of thermal injury. V. The predictability and the significance of thermally induced rate processes leading to irreversible epidermal injury', Arch. Pathol. **43**, 489 (1947).
[34] J. Pearce and S. Thomsen, 'Rate process analysis of thermal damage', in *Optical-Thermal Response of Laser-Irradiated Tissue*, edited by A. J. Welch and M. van Germert (Plenum Press, New York, 1995), pp. 561–606.
[35] D. M. Simanowskii, M. A. Mackanos, A. R. Irani, C. E. O'Connel-Rodwell, C. H. Contag, H. A. Schwettman, and D. V. Palanker, 'Cellular tolerance to pulsed hyperthermia', Phys. Rev. E **74**, 911 (2005).
[36] D. M. Harris, D. Fried, L. Reinisch, T. Bell, D. Schlachter, L. From, and J. Burkart, 'Eyelid resurfacing', Lasers Surg. Med. **25**, 107 (1999).
[37] I. Apitz and A. Vogel, 'Material ejection in nanosecond Er:YAG laser ablation of water, liver, and skin', Appl. Phys. A **81**, 329 (2005).
[38] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, Oxford, 1959), 2nd ed.
[39] R. R. Anderson and J. A. Parrish, 'Selective photothermolysis – precise microsurgery by selective absorption of pulsed radiation', Science **220**, 524 (1983).
[40] V. Venugopalan, N. S. Nishioka, and B. B. Mikic, 'The thermodynamic response of soft biological tissues to pulsed ultraviolet laser radiation', Biophys. J. **69**, 1259 (1995).
[41] V. Venugopalan, N. S. Nishioka, and B. B. Mikic, 'Thermodynamic response of soft

biological tissues to pulsed infrared laser radiation', Biophys. J. **70**, 2981 (1996).
[42] G. Paltauf and P. E. Dyer, 'Photomechanical processes and effects in ablation', Chem. Rev. **103**, 487 (2003).
[43] M. W. Sigrist, 'Laser generation of sound waves in liquids and gases', J. Appl. Phys. **60**, R83 (1986).
[44] J. C. Bushnell and D. J. McCloskey, 'Thermoelastic stress generation in solids', J. Appl. Phys. **39**, 5541 (1968).
[45] G. Paltauf and H. Schmidt-Kloiber, 'Microcavity dynamics during laser-induced spallation of liquids and gels', Appl. Phys. A **62**, 303 (1996).
[46] I. Itzkan, D. Albagli, M. L. Dark, L. T. Perelman, C. von Rosenberg, and M. Feld, 'The thermoelastic basis of short pulsed laser ablation of biological tissue', Proc. Natl. Acad. Sci. USA **92**, 1960 (1995).
[47] G. Paltauf and H. Schmidt-Kloiber, 'Photoacoustic cavitation in spherical and cylindrical absorbers', Appl. Phys. A **68**, 525 (1999).
[48] M. Frenz, G. Paltauf, and H. Schmidt-Kloiber, 'Laser-generated cavitation in absorbing liquid induced by acoustic diffraction', Phys. Rev. Lett. **76**, 3546 (1996).
[49] F. W. Dabby and U. C. Paek, 'High-intensity laser-induced vaporization and explosion of solid material', IEEE J. Quant. Electron. **8**, 106 (1972).
[50] A. Miotello and R. Kelly, 'Critical assessment of thermal models for laser sputtering at high fluences', Appl. Phys. Lett. **67**, 3535 (1995).
[51] R. W. Schrage, *A theoretical study of interphase mass transfer* (Columbia University Press, New York, 1953).
[52] A. D. Yablon, N. S. Nishioka, B. B. Mikic, and V. Venugopalan, 'Physical mechanisms of pulsed infrared laser ablation of biological tissues', in *Proc. SPIE vol. 3343 – High-Power Laser Ablation*, edited by C. R. Phipps (SPIE, Bellingham, 1998), pp. 69–77.
[53] V. Venugopalan, N. S. Nishioka, and B. B. Mikic, 'The effect of laser parameters on the zone of thermal injury produced by laser ablation of biological tissue', Trans. ASME J. Biomech. Eng. **116**, 62 (1994).
[54] A. Miotello and R. Kelly, 'Laser-induced phase explosion: New physical problems when a condensed phase approaches the thermodynamic critical temperature', Appl. Phys. A **69**, S67 (1999).
[55] P. Debenedetti, *Metastable Liquids: Concepts and Principles* (Princeton University Press, Princeton, NJ, 1996).
[56] V. P. Skripov, *Metastable Liquids* (Wiley, New York, 1974).
[57] V. P. Skripov, E. N. Sinitsyn, P. A. Pavlov, G. V. Ermakov, G. N. Muratov, N. V. Bulanov, and V. G. Baidakov, *Thermophysical properties of liquids in the metastable (superheated) state* (Gordon and Breach Science Publishers, New York, 1988).
[58] R. E. Apfel, 'Water superheated to 279.5 °C at atmospheric pressure', Nature Phys. Sci. **238**, 63 (1972).
[59] M. M. Martynyuk, 'Phase explosion of a metastable fluid', Combust. Explos. Shock Waves **13**, 178 (1977).
[60] I. I. Frenkel, *Kinetic theory of liquids* (Dover Publications, New York, 1955).
[61] S. B. Kiselev, 'Kinetic boundary of metastable states in superheated and stretched liquids', Physica A **269**, 252 (1999).
[62] B. Majaron, P. Plestenjak, and M. Lukac, 'Thermo-mechanical laser ablation of soft biological tissue: modelling the micro-explosions', Appl. Phys. B **69**, 71 (1999).
[63] R. M. Verdaasdonk, C. Borst, and M. J. C. van Germert, 'Explosive onset of continuous wave laser tissue ablation', Phys. Med. Biol. **35**, 1129 (1990).
[64] G. L. LeCarpentier, M. Motamedi, L. P. McMath, S. Rastegar, and A. J. Welch, 'Continuous wave laser ablation of tissue: analysis of thermal and mechanical events', IEEE

Trans. Biomed. Eng. **40**, 188 (1993).
[65] M. Frenz, V. Romano, A. D. Zweig, H. P. Weber, N. I. Chapliev, and A. V. Silenok, 'Instabilities in laser cutting of soft media', J. Appl. Phys. **66**, 4496 (1989).
[66] A. D. Zweig, 'A thermo-mechanical model for laser ablation', J. Appl. Phys. **70**, 1684 (1991).
[67] Q. Lu, 'Thermodynamic evolution of phase explosion during high-power nanosecond laser ablation', Phys. Rev. E **67** (2003).
[68] R. S. Dingus, D. R. Curran, A. A. Oraevsky, and S. L. Jacques, 'Microscopic spallation process and its potential role in laser-tissue ablation', in *Proc. SPIE vol. 2134A – Laser-Tissue Interaction V*, edited by S. L. Jacques (SPIE, Bellingham, 1994), pp. 434–445.
[69] G. Paltauf and H. Schmidt-Kloiber, 'Model study to investigate the contribution of spallation to pulsed laser ablation of tissue', Lasers Surg. Med. **16**, 277 (1995).
[70] A. A. Oraevsky, S. L. Jacques, R. O. Esenaliev, and F. K. Tittel, 'Pulsed laser ablation of soft tissue, gels, and aqueous solutions at temperatures below 100°C', Lasers Surg. Med. **18**, 231 (1996).
[71] G. E. Duvall and G. R. Fowles, 'Shock waves', in *High Pressure Physics and Chemistry*, edited by R. S. Bradley (Academic Press, New York, 1963), pp. 209–291.
[72] Y. B. Zel'dovich and Y. P. Raizer, *Physics of Shock Waves and High Temperature Hydrodynamic Phenomena*, vol. I and II (Academic Press, New York and London, 1966).
[73] Q. Lu, S. S. Mao, X. Mao, and R. E. Russo, 'Delayed phase explosion during high-power nanosecond laser ablation of silicon', Appl. Phys. Lett. **80**, 3072 (2002).
[74] A. Vogel and J. Noack, 'Shock wave energy and acoustic energy dissipation after laser-induced breakdown', in *Proc. SPIE vol. 3254 – Laser-Tissue Interaction IX*, edited by S. L. Jacques (SPIE, Bellingham, 1998), pp. 180–189.
[75] D. Albagli, L. T. Perelman, G. S. Janes, C. von Rosenberg, I. Itzkan, and M. Feld, 'Inertially confined ablation of biological tissue', Lasers Life Sci. **6**, 55 (1994).
[76] A. Vogel, I. Apitz, S. Freidank, and R. Dijkink, 'Sensitive high-resolution white-light Schlieren technique with large dynamic range for the investigation of ablation dynamics', Opt. Lett. **31**, 1812 (2006).
[77] E. Hecht and A. Zajac, *Optics* (Addison Wesley, Reading, MA, 1977).
[78] J. T. Walsh and T. F. Deutsch, 'Measurement of Er:YAG laser ablation plume dynamics', Appl. Phys. B **52**, 217 (1991).
[79] Z. Bor, B. Hopp, B. Rácz, G. Szabó, I. Ratkay, I. Süveges, A. Füst, and J. Mohay, 'Plume emission, shock wave and surface wave formation during excimer laser ablation of the cornea', Refract. Corneal Surg. (Suppl.) **9**, S111 (1993).
[80] J. P. Cummings and J. T. Walsh, 'Q-switched ablation of tissue: plume dynamics and the effect of tissue mechanical properties', in *Proc. SPIE vol. 1646 – Laser-Tissue Interaction III*, edited by S. L. Jacques (SPIE, Bellingham, 1992), pp. 242–253.
[81] R. R. Krueger and S. L. Trokel, 'Quantitation of corneal ablation by ultraviolet laser light', Arch. Ophthalmol. **103**, 1741 (1985).
[82] J. Noack, R. Tönnies, C. Hohla, R. Birngruber, and A. Vogel, 'Influence of ablation plume dynamics on the formation of central islands in excimer laser photorefractive keratectomy', Ophthalmology **104**, 823 (1997).
[83] N. Arnold, J. Gruber, and J. Heitz, 'Spherical expansion of the vapor plume into ambient gas: an analytical model', Appl. Phys. A **69**, S87 (1999).
[84] H. L. Brode, 'Blast wave from a spherical charge', Phys. Fluids **2**, 217 (1959).
[85] G. Taylor, 'The formation of a blast wave by a very intense explosion. I Theoretical discussion', Proc. Roy. Soc. A **201**, 159 (1950).

[86] L. I. Sedov, *Similarity and Dimensional Methods in Mechanics* (Academic Press, New York, 1959).
[87] L. D. Landau and E. M. Lifschitz, *Fluid Mechanics* (Pergamon Press, Oxford, 1987), 2nd ed.
[88] A. Sakurai, 'On the propagation and structure of a blast wave, I', J. Phys. Soc. Japan **8**, 662 (1953).
[89] A. Sakurai, 'On the propagation and structure of a blast wave, II', J. Phys. Soc. Japan **9**, 256 (1954).
[90] D. A. Freiwald and R. A. Axford, 'Approximate spherical blast theory including source mass', J. Appl. Phys. **46**, 1171 (1975).
[91] R. Kelly and B. Braren, 'On the direct observation of the gas dynamics of laser-pulse sputtering of polymers. Part I: Analytical considerations', Appl. Phys. B **53**, 160 (1991).
[92] R. Kelly and A. Miotello, 'Pulsed-laser sputtering of atoms and molecules. Part I: Basic solutions for gas-dynamic effects', Appl. Phys. B **57**, 145 (1993).
[93] P. E. Dyer and J. Sidhu, 'Spectroscopic and fast photographic studies of excimer laser polymer ablation', J. Appl. Phys. **64**, 4657 (1988).
[94] D. L. Jones, 'Intermediate strength blast wave', Phys. Fluids **11**, 1664 (1968).
[95] C. Stauter, P. Grard, J. Fontaine, and T. Engel, 'Laser ablation acoustical monitoring', Appl. Surf. Sci. **109/110**, 174 (1997).
[96] D. A. Freiwald, 'Approximate blast wave theory and experimental data for shock trajectories in linear explosive-driven shock tubes', J. Appl. Phys. **43**, 2224 (1972).
[97] C. E. Otis, B. Braren, M. O. Thompson, D. Brunco, and P. M. Goodwin, 'Mechanisms of excimer laser ablation of strongly absorbing systems', in *Proc. SPIE vol. 1856 – Laser Radiation Photophysics*, edited by B. Braren and M. N. Libenson (SPIE, Bellingham, 1993), pp. 132–142.
[98] Z. Chen, A. Bogaerts, and A. Vertes, 'Phase explosion in atmospheric pressure infrared laser ablation from water-rich targets', Appl. Phys. Lett. **89** (2006).
[99] K. Nahen and A. Vogel, 'Shielding by the ablation plume during Er:YAG Laser ablation', J. Biomed. Opt. **7**, 165 (2002).
[100] K. Nahen and A. Vogel, 'Investigations on acoustic on-line monitoring of IR laser ablation of burned skin', Lasers Surg. Med. **25**, 69 (1999).
[101] K. Nahen and A. Vogel, 'Acoustic signal characteristics during IR laser ablation, and their consequences for acoustic tissue discrimination', in *Proc. SPIE vol. 3914 – Laser-Tissue Interaction XI: Photochemical, Photothermal, and Photomechanical*, edited by D. D. Duncan, J. O. Hollinger D. D. S., and S. L. Jacques (SPIE, Bellingham, 2000), pp. 166–176.
[102] A. Vogel, B. Kersten, and I. Apitz, 'Material ejection in free-running Er:YAG laser ablation of water, liver and skin by phase explosion, confined boiling, recoil-induced expulsion and flow-induced suction', in *Proc. SPIE vol. 4961 – Laser-Tissue Interaction XIV*, edited by S. L. Jacques, D. D. Duncan, S. J. Kirkpatrick, and A. Kriete (SPIE, Bellingham, 2003), pp. 40–47.
[103] R. Hibst, *Technik, Wirkungsweise und medizinische Anwendungen von Holmium- und Erbium-Lasern* (Ecomed, Landsberg, 1996).
[104] C. A. Puliafito, D. Stern, R. R. Krueger, and E. R. Mandel, 'High-speed photography of excimer laser ablation of the cornea', Arch. Ophthalmol. **105**, 1255 (1987).
[105] D. Bäuerle, *Laser Processing and Chemistry* (Springer, Berlin, 2000).
[106] J. P. Cummings and J. T. Walsh, 'Tissue tearing caused by pulsed laser-induced ablation pressure', Appl. Opt. **32**, 494 (1993).
[107] C. R. Phipps, R. F. Harrison, T. Shimada, G. W. York, T. P. Turner, X. F. Corlis,

H. S. Steele, L. C. Haynes, and T. R. King, 'Enhanced vacuum laser-impulse coupling by volume absorption at infrared wavelengths', Lasers Particle Beams **8**, 281 (1990).
[108] P. E. Dyer and R. K. Al-Dhahir, 'Transient photoacoustic studies of laser tissue ablation', in *Proc SPIE vol. 1202 – Laser Tissue Interaction*, edited by S. L. Jacques (SPIE, Bellingham, 1990), pp. 46–60.
[109] F. Könz, M. Frenz, H. Pratisto, H. P. Weber, H. Lubatschowski, O. Kermani, W. Ertmer, H. J. Altermatt, and T. Schaffner, 'Thermal and mechanical damage of corneal tissue after free-running and Q-switched mid-infrared laser ablation', in *Proc. SPIE vol 2077 – Laser Interaction with Hard and Soft Tissue*, edited by M. J. C. van Gemert, R. W. Steiner, L. O. Svaasand, and H. Albrecht (SPIE, Bellingham, 1994), pp. 78–86.
[110] R. O. Esenaliev, A. A. Oraevsky, V. S. Letokhov, A. A. Karabutov, and T. V. Malinsky, 'Studies of acoustical and shock waves in the pulsed laser ablation of biotissue', Lasers in Surg. Med. **13**, 470 (1993).
[111] O. G. J. Engel, 'Crater depth in fluid impacts', J. Appl. Phys. **37**, 1798 (1966).
[112] A. Prosperetti and H. N. Oguz, 'The impact of drops on liquid surfaces and the underwater noise of rain', Annu. Rev. Fluid Mech. **25**, 577 (1993).
[113] J. T. Walsh and T. F. Deutsch, 'Er:YAG laser ablation of tissue: Measurement of ablation rates', Lasers Surg. Med. **9**, 327 (1989).
[114] S. Watanabe, T. J. Flotte, D. J. McAuliffe, and S. L. Jacques, 'Putative photoacoustic damage in skin induced by pulsed ArF excimer laser', J. Invest. Dermatol. **90**, 761 (1988).
[115] Y. Yashima, D. J. McAuliffe, S. L. Jacques, and T. J. Flotte, 'Laser-induced photoacoustic injury of skin: effect of inertial confinement', Lasers Surg. Med. **11**, 62 (1991).
[116] R. Pini, F. Rossi, S. Salimbeni, S. Siano, M. Vannini, F. Carones, G. Trabucci, R. Brancato, and P. G. Gobbi, 'Experimental investigation on acoustic phenomena induced inside the eyeball by excimer laser ablation of the cornea', in *Proc. SPIE vol. 2632 – Lasers in Ophthalmology III*, edited by R. Birngruber and A. F. Fercher (SPIE, Bellingham, 1966), pp. 25–29.
[117] N. S. Nishioka and Y. Domankevitz, 'Reflectance during pulsed holmium laser irradiation of tissue', Lasers Surg. Med. **9**, 375 (1989).
[118] R. Kaufmann and R. Hibst, 'Pulsed erbium:YAG laser ablation in cutaneous surgery', Lasers Surg. Med. **19**, 324 (1996).
[119] V. Venugopalan, N. S. Nishioka, and B. B. Mikic, 'The effect of CO_2 laser pulse repetition rate on tissue ablation rate and thermal damage', IEEE Trans Biomed. Eng. **38**, 1049 (1991).
[120] J. T. Walsh and T. F. Deutsch, 'Pulsed CO_2 laser tissue ablation: Measurement of the ablation rate', Lasers Surg. Med. **8**, 264 (1988).
[121] J. K. Shimmick, W. B. Telfair, C. R. Munnerlyn, J. D. Bartlett, and S. L. Trokel, 'Corneal ablation profilometry and steep central islands', J. Refract. Surg. **13**, 235 (1997).
[122] L. V. Zhigilei and B. J. Garrison, 'Microscopic mechanisms of laser ablation of organic solids in the thermal and stress confinement irradiation regimes', J. Appl. Phys. **88**, 1281 (2000).
[123] L. V. Zhigilei, E. Leveugle, B. J. Garrison, Y. G. Yingling, and M. I. Zeifman, 'Computer simulations of laser ablation of molecular substrates', Chem. Rev. **103**, 321 (2003).
[124] R. Knochenmuss and L. V. Zhigilei, 'Molecular dynamics model of Ultraviolet matrix-assisted laser desorption/ionization including ionization process', J. Phys. Chem. B **109**, 22947 (2005).
[125] K. T. Schomacker, J. T. Walsh, T. J. Flotte, and T. F. Deutsch, 'Thermal damage

produced by high-irradiance continuous wave CO_2 laser cutting of tissue', Lasers Surg. Med. **10**, 74 (1990).

[126] M. Frenz, H. Pratisto, F. Könz, E. D. Jansen, A. J. Welch, and H. P. Weber, 'Comparison of the effects of absorption coefficient and pulse duration of 2.12 μm and 2.79 μm radiation on laser ablation of tissue', IEEE J. Quantum Electron. **32**, 2025 (1996).

[127] A. Vogel, P. Schmidt, and B. Flucke, 'Minimization of thermomechanical side effects in IR ablation by use of multiply Q-switched laser pulses', Med. Laser Appl. **17**, 15 (2002).

Oscillations, Waves and Interactions, pp. 259–278
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-10-8

Laser speckle metrology – a tool serving the conservation of cultural heritage

K. D. Hinsch

Applied Optics, Institute of Physics, Carl von Ossietzky University Oldenburg
26111 Oldenburg, Germany

Abstract. Deterioration of artwork is often connected to mechanical material degradation that starts at microscopic scales. Insight into decay mechanisms can therefore be obtained by monitoring microscopic deformation and displacement fields. Thus, the proper optical methods become an ideal tool for restorers and conservators, the more as the methods are non-intrusive and remotely applicable. We show how the scope of modern speckle metrology can be adapted to this aim. Refinements of correlation imaging, speckle interferometry and low-coherence detection as well as the time average monitoring of vibrations provide a wealth of methods that have been applied successfully in historical objects.

1 Introduction

For many years the preservation of artwork was mainly the domain of the humanities. Lately, however, science is playing an increasing role in the analysis, conservation and restoration of historical objects. For unique and delicate specimens – from skilfully carved stone sculptures to colourful medieval wall paintings – analytical techniques must be non-invasive. Optics is providing a powerful and versatile set of tools for surveying and analyzing historical treasures [1–3]. Since they were made these priceless objects have aged. They deteriorate due to varying climate conditions or polluted environments and their preservation requires special countermeasures. Critical monitoring of the state of the object, identification and understanding of the deterioration processes and control of remedies are important.

Often, deterioration starts at the microscopic level, initially producing weakening of the mechanical cohesion in the sample. This shows up in irregular minute displacements or changes in the micro-topography of an objects surface. Thus, optical contouring and deformation mapping methods can provide essential data on the distribution of mechanical stress in the sample, indicate weak spots, or provide early-warning data on objects at risk. Suitable methods must provide sufficient sensitivity to detect displacements or changes in the topography well down in the micrometer domain. Yet, they should still operate successfully on-site in spite of disturbances like rigid-body creeping motions, annoying vibrations or air turbulence. The state of the object may also be checked by monitoring its deformation response to an external mechanical or thermal load (comparable to the medical checkup in health care).

Generally, any of the optical methods for displacement mapping that are employed in experimental mechanics are also suited for the present task – provided they are robust enough to be applied in an environment outside a laboratory. Optical techniques are often based on the evaluation of the laser light field scattered from the surface under inspection. This field is characterized by its random nature – manifest in the speckled appearance of the image of the object. This laser speckle pattern can be considered a "fingerprint" of the surface and its analysis may provide the wanted data on object changes. This may be done by correlation of the speckle intensity fields using comparatively simple equipment. The motion of the fingerprint, so to speak, provides displacement data at sensitivity well in the micrometer range. Thus, the delicate response of pieces of art to environmental loads or an artificial external stimulus can be monitored. At the same time any small change in the fingerprint pattern provides a measure for average changes in the surface profile that may quantify the deterioration attack on the surface. For the detection of still smaller displacements (sub-wavelength sensitivity) interferometric methods that measure phase shifts are needed.

The present article introduces characteristic features of speckle metrology and illustrates its basic performance in art monitoring. Then it concentrates on recent sophisticated refinements to extend the performance to specific situations at delicate objects or in unfavourable measurement environments that otherwise would rule out this kind of sensitive optical metrology. Finally, we devise novel approaches that explore the situation even underneath a surface.

2 Digital image correlation – displacement fields, surface deterioration and ablation monitoring

When a rough-surface object is viewed or photographed in laser-light illumination its image is covered with a granular pattern of speckles [4]. These are produced by interference of the many elementary light waves scattered from the irregularities of the surface. The random phases imposed on the light produce interference of statistically varying outcome. Thus the speckle pattern is taken as the fingerprint of the surface – encoded by the optical system doing the observation. Any displacement of the surface will produce an according motion of the speckle pattern and any changes in the microscopic topography of the surface will alter the speckle pattern.

To measure the displacement field of a sample surface we just need to take an image of the object before and after the motion – usually with an electronic camera. The images are then evaluated by determining locally the shifts of the recorded speckle patterns. For this purpose, the images are subdivided into a matrix of small interrogation regions. For each of the sub-images the displacement field is computed from a two-dimensional cross correlation – thus the technique is often termed Digital Speckle Correlation or Digital Image Correlation (DIC) [5]. The advantage over traditional techniques is that the surface need not show visual details nor do we have to affix any markings to the object. Furthermore, high sensitivity can be obtained by using small-sized speckles which is achieved by stopping down the imaging aperture. Sophisticated algorithms yield the displacement field with sub-pixel resolution

of roughly a micrometer. Mind, however, that the method is primarily limited to detect displacements normal to the viewing direction, the so-called in-plane component. The experimental setup is extremely simple consisting of light source, CCD-camera and computer.

Practically, the displacement value is obtained from the position of the cross-correlation peak. Successful performance of the method thus requires a certain similarity of the speckle patterns in both the images compared – otherwise correlation is lost and the peak degrades. When decorrelation of the speckle patterns becomes an issue during large displacements or significant changes in the surface a white-light version of image correlation may help. Provided there is sufficient image texture from details in the object the laser is replaced by a traditional light source and correlation is based on the motion of image details. The resolution in this case is set by the fineness of the image details.

DIC in either version is an established technique and has been applied repeatedly also to tasks in artwork diagnostics. For on-site investigations the white-light version provides a robust setup of sensitivity well in the μm-range. We used it, for example, in the monitoring of the response of antique leather tapestry to changes in temperature and humidity. Laser speckle instruments were used in the same project to determine basic data of leather under mechanical and thermal loads [6].

In its primary form, DIC gives only the in-plane displacement data. It was shown, however, that it can even provide a three-dimensional displacement vector. An analysis of the shape of the correlation peak or the cross power spectrum of the speckle images under comparison provides the local tilt from which the out-of-plane displacement can be calculated by integration [7]. The accuracy of the component thus obtained, however, falls one order of magnitude short of the in-plane component.

We mentioned that decorrelation is of disadvantage in displacement mapping. On the other hand, the reduction in the correlation coefficient can be used as a measure for changes in the topography of a rough surface. These could be evidence for micro-scale processes fuelling artwork deterioration. We have employed such decorrelation analysis for various issues in artwork research [8]. Typical situations are the monitoring of salt crystals growing on historical murals or an estimate of the impact of repeated water condensation on the historical substance in natural building stones. Let us show the typical performance by another example.

Lasers are often used to clean artwork by ablation of dirt. Efficient monitoring of the ablation process is needed to make sure that only the dirt is removed and no damage is done to the invaluable substance underneath. For this purpose we took speckle images of the surface under treatment for each laser shot and used the decrease in the correlation coefficient to indicate the amount of matter removed. A thorough analysis with varying surface models had to be carried out to quantify the relation between average change in the surface profile and decrease in correlation [8]. Fig. 1 presents a set of correlation coefficients versus the number of laser shots in the cleaning of a sandstone sample by green 400-mJ Nd:YAG-laser pulses. Our modelling allowed us to assign an average removal of 70 nm to a correlation coefficient of about 0.1. The family of graphs is obtained by taking every second image as a new reference image. The results show that the first pulse already removes about 70 nm; later pulses, however, produce decreasing effects. In suitable object/dirt combinations we

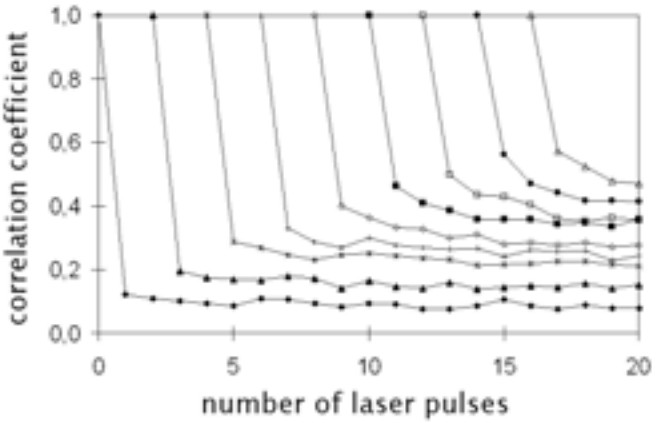


Figure 1. Monitoring of dirt removal from a historic sandstone sample by speckle correlation. Material is ablated by green Nd:YAG-laser pulses. The family of graphs is obtained by using new reference images in the course of the process.

expect that the rate of ablation will change markedly when the underlying substance is reached.

3 Video holography – mechanical response of historical murals to sunshine

Displacements well below a micron can be detected by interferometric methods. This is achieved by phase-sensitive recording of the speckle light with the aid of a reference wave in a setup for video-holography, mostly called Electronic Speckle Pattern Interferometry (ESPI) [9]. Here, too, the object under investigation is illuminated by laser light and imaged by a CCD-camera (Fig. 2). Now, however, a spherical reference wave is superimposed via a beamsplitter producing an interference image (image plane hologram) available for further processing in a computer. Subtraction of successive images, for example, yields a system of so-called correlation fringes. These are contour lines of constant displacement in the direction of a sensitivity vector $\boldsymbol{k}$ that is determined by the difference in illumination and observation directions. For normal illumination and viewing, for example, $\boldsymbol{k}$ points into the out-of-plane direction. In combining several optical setups of different geometry all three spatial components of a displacement vector can thus be determined. The interpretation of a single fringe system is ambiguous – it does not tell us the sign of the displacement. Furthermore, jumps in the fringes that occur when discontinuities in the displacement field are involved impede evaluation. To handle these problems several images need to be recorded for each of which the reference wave has undergone a set phase shift. Often three or four images at regular phase intervals are collected – a strategy

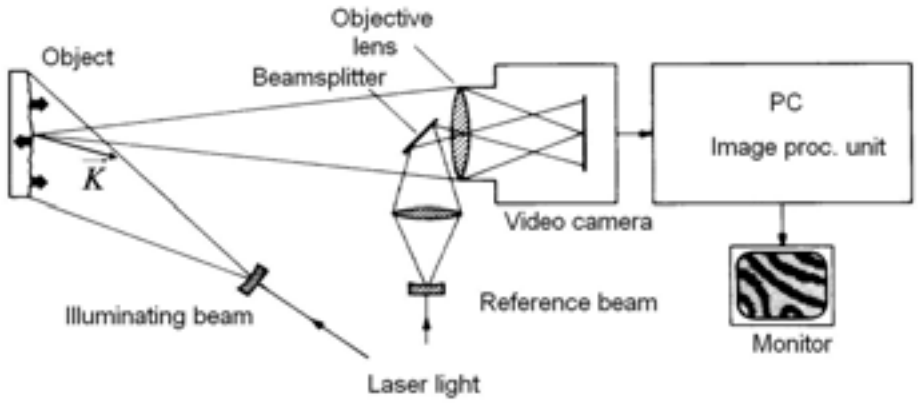


Figure 2. Optical setup for deformation measurement by video holography (Electronic Speckle Pattern Interferometry – ESPI).

known from classical interferometry as phase shifting. Such a series of images allows automated evaluation by a procedure called spatial phase unwrapping and provides unambiguous deformation data [10].

ESPI may also be used in the measurement of amplitudes in sinusoidal vibrations – a feature that we will use in the monitoring of loose plaster layers at murals. If we assume that the vibration period is short compared with the recording time the signal in the electronic camera is a time average over the interference signal at all phases. Proper electronic or digital processing of this signal produces fringes contouring the vibration amplitude.

To introduce the matter, outline the problems that need improvements, and illustrate the type of data obtained we show some early results from a project where traditional ESPI provided answers to urgent questions of the conservators. The work concerned the role of solar irradiation in the decay of 19th-century murals in Wartburg castle of Thuringia, Germany. This castle is the famous place where Martin Luther translated the bible into German around 1520. The legend says that he had to fend off the devil by throwing his inkpot at him! A conventional ESPI system was rigidly attached to the wall. It used laser-diode illumination and phase shifting by successive piezo-electrically driven tilt of a glass plate in the reference wave. The thermal load on the fresco was estimated by mapping deformations during cyclic heating and cooling – sunshine being simulated by infrared irradiation. In the left part of Fig. 3 we show the area of observation in the mural and at the right the deformation field produced during a 2.5-minute period in a cooling-off phase. A characteristic feature is a sudden kink in the displacement running through the field and coinciding with an image detail in the painting. Obviously, such an abrupt change will be accompanied by high local tensions in the material that will pose a threat to the integrity of the substance – a good reason to ban all direct sunlight from the frescos.

The coincidence of the location of the displacement irregularity with the feature line in the painting provided an interesting explanation for the discontinuity in the

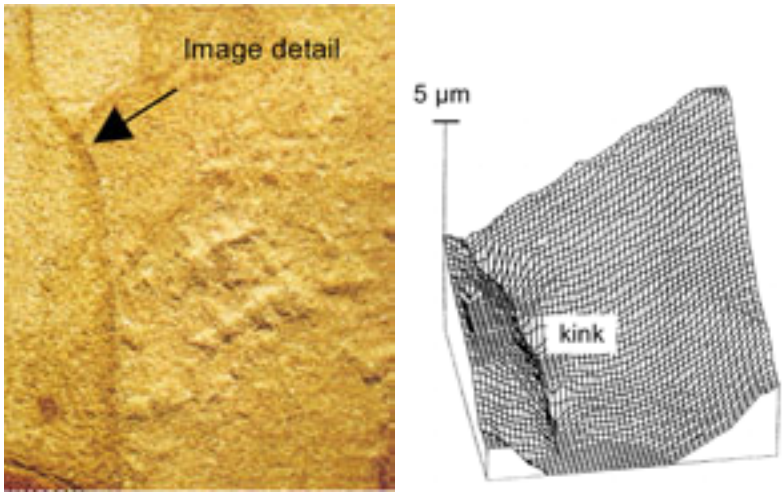


Figure 3. ESPI study of the thermal response of a fresco in Wartburg castle, Thuringia, Germany. Left: measuring field (width 6 cm); right: displacement map showing deformation kink at the location of a former boundary of work-piece.

mechanical response. Here, two work pieces meet that the painter has produced at different times. Frescos are painted onto fresh plaster and the artist scrapes off unused plaster when he finishes a days work. Our measurement result suggests that new plaster does not attach well to the old. The poor mechanical contact is a source for future problems. 150 years after the frescos were painted we uncovered where the artist made a break in his work – although he tried to hide it underneath the feature line in the image!

This example illustrates a typical implementation of optical metrology in the monitoring of artwork. Long-term changes in a specimen are difficult to register directly because this would require a stable measuring device at the object for a very long time. Rather, such changes are revealed in regular checkups by repeatedly exposing the specimen to a standard load and studying any changes in its response.

4 New challenges for ESPI (electronic speckle pattern interferometry)

Our continuing optical activities at cultural-heritage objects revealed that early-day ESPI needed substantial improvements to cope with the problems generally met in the practical study of artwork. The quality of the measurements suffered from spurious signal fluctuations during observation and data acquisition that originated from background vibrations, turbulence in the optical path or rigid-body misalignments. Sometimes, the technique was just too sensitive for the process encountered, sometimes the deformation rate was too rapid to obtain correct phase-shifted data; often the light source lacked coherence or was instable. Let us therefore turn towards more refined techniques that take up such shortcomings and provide some novel innovative approaches.

We have learned that a successful ESPI system requires provisions for phase shifting. For each state of the object several frames are needed that have been taken at a set of given phase differences between object and reference wave. The system used at the Wartburg employed temporal phase shifting (TPS) where the phase shifts are produced in succession by changing the optical path length in one optical branch. This can be imposed by turning a glass plate, translating a mirror or stressing a glass fibre in the optical delivery. Processing of the phase-shifted frames produces a phase map mod 2π which is displayed in a saw tooth grey-level or pseudo-colour representation.

During live observations of non-stationary objects like pieces of art in their everyday environment the conditions may change in time faster than allowed by the time-out required for phase shifting. Even if the object deformation is sufficiently slow, the measured phase is often deteriorated by air turbulences in the optical path or by background vibrations. Therefore, schemes have been developed to obtain the phase data simultaneously.

The most successful concept in which at least three phase-shifted images are recorded on the same CCD-target in the camera is called spatial phase shifting (SPS) [11,12]. For this purpose the source of the spherical reference wave originating in the aperture of the imaging optics (cf. Fig. 2) is given a small lateral offset resulting in a linear increase of the phase along one direction on the target. This offset must be adjusted such that the period of the carrier fringes resulting from interference of object and reference wave equals three times the pixel pitch in the offset direction. Then the reference-wave phase between adjacent pixels differs by $120\,^\circ$ and the combination of data from three neighbouring pixels each gives the necessary phase-shifted frames. These can be combined in the commonly used phase-shifting algorithms to produce the mod 2π saw tooth pattern. For our purposes another evaluation method, the Fourier-transform technique [13], is more appropriate, because it offers additional features as we will see soon.

Let us briefly describe this technique which yields the complex-valued (amplitude and phase) light distribution in the image plane that can be utilized to calculate the phase difference data of subsequent images needed to determine the displacement. We mentioned the carrier fringes due to the superposition of object-light and off-axis reference wave. Fourier-transformation of the CCD-image thus yields a zero-order term and two side-band terms at the carrier frequency, both of which contain the complex spatial frequency spectrum of the object light. Careful matching of pixel data (size and pitch), speckle size and reference wave offset will guarantee non-overlapping spectral terms and maximum free spectral range. Thus, any one of the sideband terms can be separated for inverse Fourier-transformation to yield the complex object-light distribution and thus the wanted object-light phase.

Let us demonstrate the superiority of spatial (SPS) over temporal (TPS) phase shifting by observing a static deformation (point-like load at the centre of a membrane) that is disturbed by background vibration (insufficient vibration insulation of the setup) or hot turbulent air in the viewing path (Fig. 4). Obviously, TPS (upper frames in Fig. 4) results in poor-quality saw tooth images when vibrations cause wrong phase shifts and thus a loss of directional information (centre) or when small-scale turbulence creates even locally varying phase shifts so that the correct-

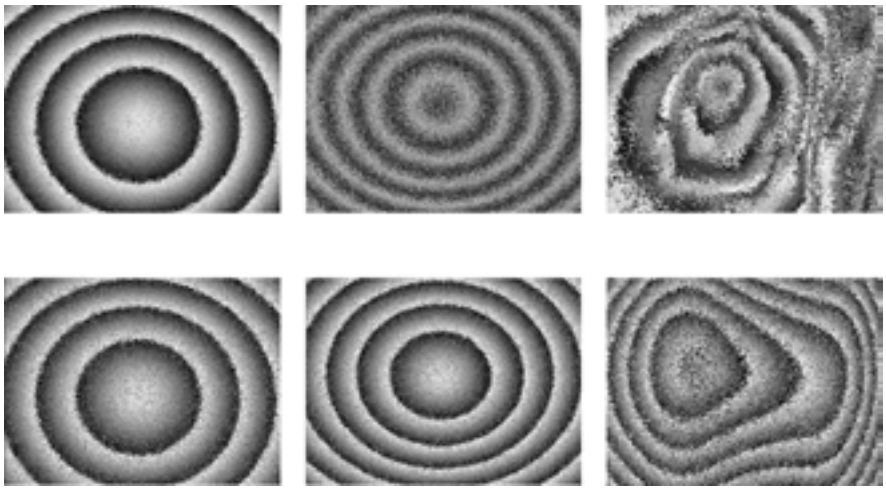


Figure 4. Deterioration of the saw tooth phase maps in an ESPI deformation study under stable experimental conditions (left) and disturbed by vibrations (centre) or air turbulence (right). Performance of temporal phase shifting TPS (upper) versus spatial phase shifting SPS (lower).

ness of the image is partly lost (right). Due to the greatly improved performance of SPS (lower frames – under the same conditions) we decided therefore to implement this arrangement whenever possible. Comparison of both the static images (left), however, illustrates that under stable conditions SPS performs slightly inferior due to residual fluctuating speckle phase over the three pixels compared [14].

For an overall displacement map characterizing the process under investigation the steps in the mod 2π-maps have to be eliminated – a procedure called spatial phase unwrapping. The usual way is to detect locations of the 2π jumps by comparing neighbouring phase values and converting the step function into a continuous displacement phase by adding the required integer multiples of 2π. This procedure is carried out along suitable tracks in the image – thus the name spatial phase unwrapping. In the history of ESPI, effective unwrapping algorithms that are resistant to a propagation of errors have been an important issue. They easily work in good-quality data fields. Real-world objects like those we are concerned with in this article, however, pose real challenges.

The problem can be nicely illustrated by the response of a piece of historic brick, 2 cm in thickness, to cyclic heating and cooling with an infrared radiator (Fig. 5). The white-light image of the specimen, Fig. 5 left, already implies some difficulties to expect. We see that a network of cracks divides the brick into numerous sub-areas that probably each will execute an independent deformation. Indeed, the out-of-plane displacement phase map mod 2π in Fig. 5 (centre) reveals several areas of irregular boundaries that have undergone separate motion as indicated by the varying fringe densities and orientations. For spatial unwrapping the areas would need identification and separate evaluation – quite a laborious task. Even more, the saw tooth image

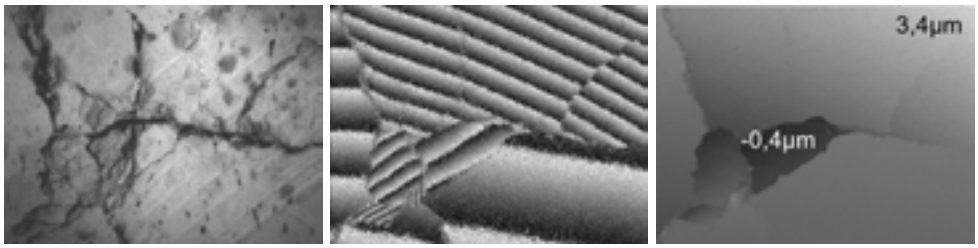


Figure 5. Mapping of out-of-plane displacements in a historic brick specimen due to heat irradiation. Left: white-light image; centre: mod 2π phase map; right: displacement obtained by temporal phase unwrapping.

does not render any information as to the relative heights in the sub-areas because the absolute fringe order within each area is not known – we have no information about the fringe count during the period between capturing images. Yet, this is an important quantity in estimating the distribution of mechanical loads between sub-areas in the specimen.

A solution to this problem is given by temporal phase unwrapping [15]. It makes use of the rapid data rate available in recent image acquisition and processing equipment. The phase history at every single pixel in the camera is stored – in combination with SPS we now get saw tooth data versus time. When images are stored at a rate excluding intermediate phase changes of more than $\pm\pi$ no ambiguous phase jumps are encountered. Thus, we are not bothered by any irregularities in space and unwrapping even yields the relative displacements between the sub-areas. Let us prove this by looking at the result in Fig. 5 right where the amount of displacement is encoded in grey level – ranging from -0.4 μm at the darkest to +3.4 μm at the brightest. In such cases, an important question by restorers is whether the deformation in the specimen is reversible and the object follows the cyclic load by cyclic motion. This could, of course, be answered only on the basis of such absolute displacement data as we have obtained them here.

The speed of a process in long-term investigations may differ considerably, e. g., when it is driven by the ambient climate as in many studies on historical objects. In this case, the rate of image storage should be adjusted dynamically. It must be high enough to obey the sampling theorem and as low as possible to save storage space. Once more, temporal phase unwrapping offers a solution, because it provides an instant fringe count. We have used this to trigger the instant of recording images for an optimum number of fringes over the viewing field [16].

Deformation measurements by ESPI rely on the local correlation of the speckle fields scattered from the object surface at different instants of time and are impeded severely when the fields decorrelate. There are two main causes for such effects. The speckle field is altered by the overall motion of the object (geometric decorrelation) or it may change by the minute changes in the surface texture already covered earlier [17]. Either kind of decorrelation will spoil the quality of the measurement and limit the range of applicability.

Often, in-plane motion causes a displacement of the small interrogation areas that are compared in the correlation. A straight-forward calculation with fixed coordinates suffers from a mismatch because only part of the data within each area contributes to correlation. This is especially pronounced in microscopic ESPI where areas of less than a square millimetre are investigated. We had to develop means to cope with this kind of geometric decorrelation. The Fourier-transform technique of handling spatially phase-shifted data provides an elegant way to do so [18].

Recall that the Fourier-transform method re-establishes the complex object-light distribution in the image plane from which we have so far used only the phase data. Yet, we can compute also the intensity distributions that are ordinary speckle images that we would get without the reference wave. Now, we can compare these images by digital image correlation and obtain the in-plane displacement field yielding the mismatch for each interrogation sub-area. The resulting values are used to backshift one of the images for better superposition and then process the ESPI data on optimal matching sub-areas. This method is known as adaptive windowing.

The improvement in the performance thus obtained is best illustrated by examples from microscopic ESPI. In the investigation of stone deterioration, for example, researchers strive to understand how pressure from the crystallization of salts within the porous stone contributes to the weakening of the material. With pore-sizes of typically less than 100 μm deformation measurements request high spatial resolution which can only be obtained under a microscope. For this purpose we have integrated a commercial microscope of long viewing distance into an ESPI setup. Another application is in the study of crack formation in historical paint layers. The famous Chinese terracotta warriors of Lin Tong, for example, loose their invaluable paint cover almost the moment they are excavated, because the originally moist paint layers break up when getting dry. We participate in testing remedies for their conservation.

Fig. 6 was obtained in an ordinary ESPI-study of paint layers on terracotta samples while the humidity was changed – a measuring field of size $230\times230\,\mu\mathrm{m}^2$ is inspected. The saw tooth pattern in Fig. 6 (left), obtained in the traditional way, shows useful

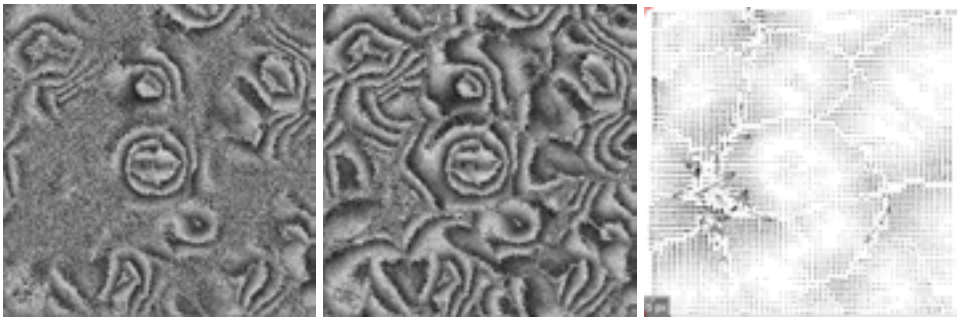


Figure 6. Elimination of geometric decorrelation in microscopic ESPI by adaptive windowing. Object ($230\times230\,\mu\mathrm{m}^2$): painted terracotta (Chinese terracotta army) under the influence of humidity. Left: original saw tooth pattern; centre: pattern improved by backshifted window; right: in-plane deformation vectors.

saw tooth fringes over certain areas, but contains several noisy regions void of fringes because the underlying speckle fields decorrelate. With back-shifting according to the strategy explained above the same data yielded the pattern of Fig. 6 (centre) showing fringes also over most of the area that could not be evaluated before. Obviously, the backshift data also give the in-plane displacement values – indicated by small arrows – supplementing the out-of-plane data from ESPI (Fig. 6, right). It is clearly seen how patches in the image that each can be attributed to a flake of paint move individually. A few spurious displacement vectors are measurement errors and would be eliminated by post-processing. Thus, a single ESPI record can now provide the complete 3D displacement field – a task that usually needs three optical configurations of complementing sensitivity vectors. In case of little surface decorrelation the in-plane component can be obtained with accuracy similar to the out-of-plane component [18].

5 Explorations into the depth: low-coherence speckle interferometry

The optical tools introduced so far make use of light that has been scattered by the specimen and carries mainly information about the location and micro-topography of the surface of a sample. Any conclusions about what is happening in the bulk of the object are indirect. Yet, many practical problems grow underneath the surface of an object and it would be of advantage to have direct access to these regions. As a typical example, take the detachment of paint and varnish layers in the Chinese terracotta-army warriors already mentioned. It is assumed that this stratified heterogeneous compound structure suffers damage because the various layers differ in their mechanical response to the change in ambient humidity. When excavated from the humid soil and moved into dry air, for example, a paint layer may shrink differently from a primary coating or the carrier material. To test such assumptions and provide for countermeasures by conservation agents it would be ideal to map deformations also for various depths in the material.

Light can be used for this purpose if it penetrates deep enough into the material and is sufficiently scattered backwards for detection. When thin paint or varnish layers are involved these are typically only some 100 μm or less in thickness. Often, there is a sufficient amount of light returning from these depths – especially from interfaces – that can be used for metrological purposes. However, a method is needed to discriminate the light according to the depth where it has been scattered. This challenge can be met by making proper use of the coherence of light as in optical coherence tomography (OCT). We have proposed a combination of OCT with ESPI that we call low-coherence speckle interferometry LCSI [19].

The basic strategy is easily explained. In ESPI, object and reference wave that are obtained by beam-splitting must superimpose coherently to preserve phase information in the CCD-images. Interference of light, however, is only possible when the optical paths travelled by the waves do not differ by more than the coherence length. Usually, researchers avoid worrying about this requirement by using laser-light sources with a coherence length exceeding all scales involved. On the other hand, low-coherence light – as from a super luminescence diode (SLD) of some 50 μm

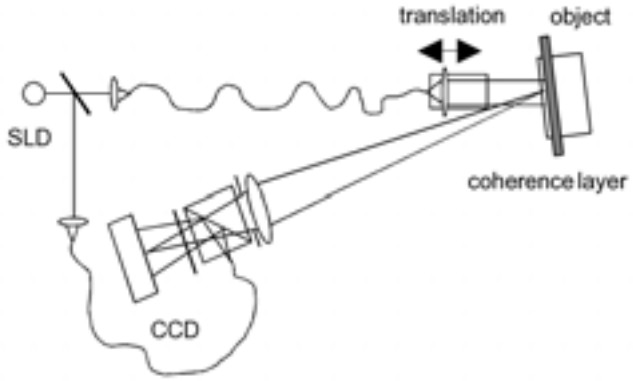


Figure 7. Optical setup for low-coherence speckle interferometry LCSI.

in coherence length – imposes restrictions on interference experiments. With this light source in combination with a proper design of the optics interference patterns can arise only for signals that differ in path accordingly little and thus identify light from a well defined layer in depth.

Fig. 7 explains the optical setup for LCSI. Light from the SLD is coupled into fibres for reference beam and object illumination. While the reference beam is fed into the ESPI optics as usual, the object light exiting from the fibre is collimated by a lens on a translation stage to illuminate the object. The geometric arrangement in the setup defines a thin "coherence layer" in space such that only light scattered within this region contributes to the interferograms evaluated for displacement. Adjustment of position and orientation of the object allows placing this layer in the desired position within the object – at an interface between two paint layers, for example. With the translation stage the coherence layer can then be scanned through the sample for investigations at varying depths.

In practice, the useful interference signal has to compete with a large amount of background light. Furthermore, the useful light has to travel in part through a complex scattering medium that even changes between the instants of observation due to the displacements in the object. Let us, for example, adjust the coherence layer onto an interface between paint and terracotta in a fragment of a Chinese warrior. We are interested in any motion of the interface during drying. On its way to and from the interface, however, the light has to pass through the bulk of paint – a path that quite probably will be influenced during drying. This produces uncorrelated changes in the light and reduces the quality of the resulting fringes. Therefore, the technique will find its limits at a certain depth that we are exploring presently.

For an estimate of the depth range available for exploration a model sample has been studied (Fig. 8) [20]. It was prepared from a partly transparent adhesive (index of refraction $n \approx 2$) that is used in bonding aluminium compounds and that was coated in steps of varying thickness onto a glass plate. The coherence layer was adjusted to the interface between adhesive and glass to observe its out-of-plane motion due to a slight tilt of the specimen. Thus, we observe evenly spaced saw tooth fringes

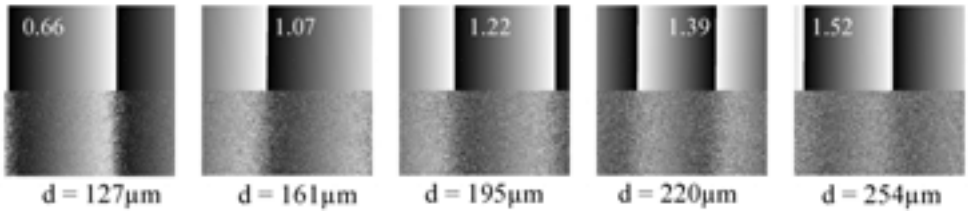


Figure 8. Performance of LCSI through scattering layers of adhesive of different thickness d. Tilt-induced saw tooth fringes for out-of-plane motion of the interface between a glass carrier and the adhesive layer. Lower row: experimental saw tooth fringes; upper row: theoretical fit yielding standard deviations given in rad.

that are parallel to the tilt axis and represent the displacement introduced. With increasing thickness of the layer of adhesive from 127 μm (Fig. 8, left) to 254 μm (Fig. 8, right) the quality of the fringes decreases due to decorrelation of the underlying speckle signals. By fitting ideal fringe functions (shown above each result) to the experimental data the standard deviation σ of the phase in radians is calculated as a measure for fringe quality [11]. The increasing values for σ as given in each fringe set indicate the evident loss in fringe quality with increasing thickness – the maximum possible value is $\sigma = 1.8$ rad and occurs in a random noise pattern that would arise with complete decorrelation. According to our results one can expect to perform successful measurements for a layer thickness of up to a few 100 μm – subject to the specific properties of the material involved.

Let us illustrate the potential of LCSI in an application during out-of-plane measurements on paint layers of the Chinese warriors. Humidity effects were studied in a terracotta fragment that carried a layer of varnish on top of a layer of paint. Thus, in addition to the interface air/varnish at the surface a second reflecting interface was located about 100 μm below the surface. By tuning the coherence layer onto either of these interfaces we wanted to measure relative motions between these layers. Fig. 9 shows deformation fringes in the $1\times1\,\mathrm{mm}^2$ sample when the system is tuned to the surface (left) or to the interface (right). The figures give the response to a decrease in ambient relative humidity from 90% to 80%.

The results from the surface indicate clearly that the sample is divided into many small sub-areas that each react by a bowl-shaped deformation due to length changes in the layers. The data originating from the interface at depth 100 μm are more difficult to read as they are noisier – a consequence of the passage of the light through the covering layer. Yet, we see the same separation into bowls, much smaller in deformation, however. A series of deformation maps allows obtaining mean displacement values for both the layers versus measuring time during a cyclic change of the humidity (Fig. 10). We verify the larger response at the top layer – which, of course, includes also the displacement at the lower layer – and realize that there is practically no delay between the reactions of the interfaces. This might be explained by unimpeded passage of humidity through a network of minute cracks separating the bowls.

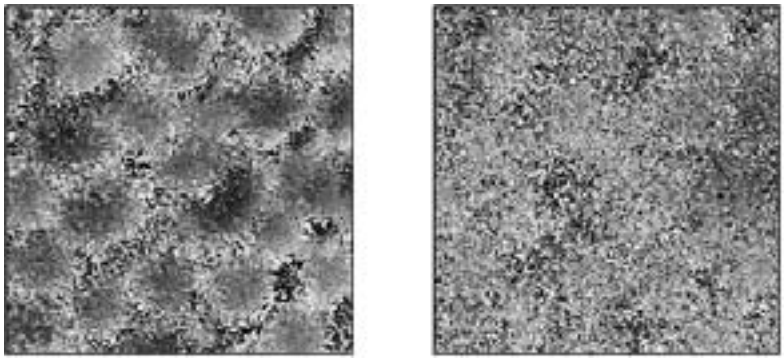


Figure 9. Humidity-induced out-of-plane motion as indicated by saw tooth fringes of the surface (left) and an inner interface at depth 100 μm (right) in a layered coating on a Chinese terracotta sample. Relative humidity changed from 90% to 80%.

In summary, LCSI is a promising tool to explore the deformation scene also within a thin region below the surface of rigid objects. The depth available depends on the optical properties of the material – mostly it will not exceed a millimetre. Since action from an aggressive environment, however, has to penetrate layers immediately adjacent to the surface their mechanical properties become especially important in estimating possible damage.

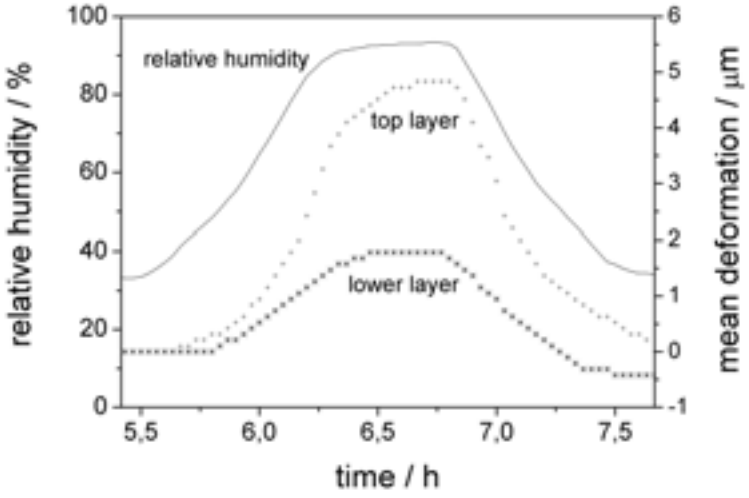


Figure 10. Mean deformation of surface (top layer) and interface (lower layer) in the layered coating on the Chinese terracotta sample of Fig. 8 during the cycle of relative humidity also shown.

6 Sounding the depth by vibration ESPI

Many historical murals are painted on plaster layers of up to several centimetres in thickness. In the course of time, such layers may detach from the supporting wall – thus a check of the integrity of the interface is needed. Conservators usually inspect the condition of wall paintings using the so-called percussion technique, which involves gently tapping the painting, section by section, and deducing from the acoustic response where the plaster is loose. At such locations it sounds "hollow". Using this method of inspection to plan the restoration of a large church is a cumbersome task requiring complex scaffolding and involving a detailed mapping of the paintings condition that can take months to complete. The obvious solution here would be to develop a measuring technique that can be used to perform this inspection quickly and automatically from ground level, allowing the experts more time to concentrate on the affected areas and their restoration.

We have shown that low-coherence exploration of an object below its surface is possible, but restricted to not much more than a millimetre in depth. Thus, in the present case light can not be used directly to scan the depth for detachments and we must develop other means. We can use the light, however, to probe the minute response of loose areas to an acoustic-wave stimulus coming from a loudspeaker and "sounding" the depth for an optical alert signal.

ESPI is a perfect tool for studying small vibrations [21] that we adapted to the mural problem [22]. In our setup (Fig. 11) an ordinary time average ESPI arrangement is refined by modulation of the reference-beam phase with a frequency slightly displaced from the loudspeaker signal. This has several advantages. In time-average interferometry the fringe function, i. e., the fringe brightness versus vibration amplitude, follows the square of a zero-order Bessel function. As the slope of this function tends to zero for zero amplitude the performance of the method is poor for small vibrations. Reference beam modulation acts as bias amplitude and allows moving the operating point in the fringe function to the place of maximum sensitivity. Furthermore, the beat between reference and signal waves produces flickering light intensities at those

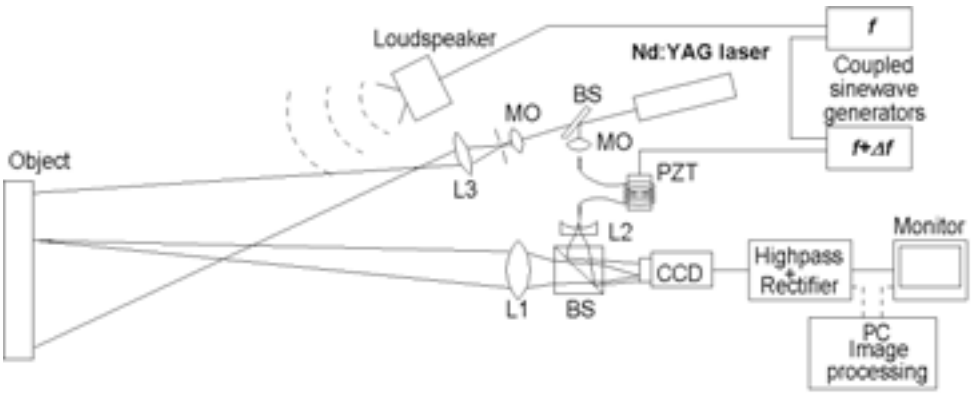


Figure 11. Time average ESPI-system for the study of detached layers in murals.

locations that take part in the vibration. This is a very intriguing feature because it implies intuitively the state of "motion" – a feature important when advertising to the community of restorers and conservators our high-tech method with difficult to read results. Finally, several images captured during a beat cycle provide the basis for temporal phase shifting to yield automated evaluation of vibration amplitude and phase. The final equipment in Fig. 11 is characterized by a fibre-optic reference link which provides the basis for phase modulation (modulation of fibre length by a PZT-driven cylinder) and, if required, allows path length matching by introducing additional fibre. The coherence constraints, however, are low since we changed from our early laser-diode illumination to a very stable CW Nd:YAG laser of many metres in coherence length. Performance is improved by robust setups and averaging over several images which allow doing on-the-site measurements of vibration amplitudes of as little as a few nanometres.

The measurement strategy in mural investigations is to look for resonances while tuning the excitation through an appropriate frequency band. For Fig. 12 the technique was tested at a specially prepared plaster layer on a stone wall that contained an artificially produced region of detachment. This was created by interrupting the mechanical contact between wall and plaster with a plastic foil. We show an amplitude map (left) and a phase map (centre) of a 230-Hz resonance mode in the loose plaster layer. Many such results obtained at 10-Hz frequency steps from 90 Hz to 580 Hz were finally accumulated to produce the final evaluation in the right of Fig. 12. The grey level indicates how often a certain location had responded to the excitation – dark areas vibrated frequently, bright ones rarely; an ordinary object image is put in the background. The location and shape of the loose region are nicely reproduced.

The new technique was put to test on frescos in a cemetery chapel at Kamenz, Saxony. The adhesion of the layers of plaster covering the walls and ceiling was examined one square metre at a time. The frequency of the sound was adjusted in steps of ten Hertz, and the reaction of the wall to each sound was recorded. For the evaluation it was again calculated how often an oscillation was detected for each position on the wall. This time the data obtained was assigned colour values, so

Figure 12. ESPI-study of vibrations of an artificially produced detachment in a plaster layer on a stone wall. Left: amplitude map of 230-Hz higher-order mode of the plaster plate; centre: corresponding phase map; right: final evaluation of excitation response between 90 Hz and 580 Hz in 10 Hz-steps indicating the loose area (dark illustrates frequently, bright rarely excited regions) – an ordinary object image in the background.

that all of the areas where there was no longer good adhesion to the substrate were eventually displayed in yellow or red. Here, the new laser-optical technique passed successfully when its results were compared to those obtained using the conventional percussion method. It benefited from the advantage that the damaged areas shown on the video image could be located precisely and automatically. Manual mapping performed by a conservator, on the other hand, can easily contain errors that creep in during the mapping process.

A famous example of where our method has been applied is the church of the Benedictine Convent of Saint John at Müstair in Graubünden, Switzerland, which has been declared a UNESCO World Heritage Site because of its medieval wall paintings. In the 12th century, the original Carolingian frescos, which had been painted 300 years before, were covered by a new layer of plaster and a series of Romanesque wall paintings. To roughen the surface in preparation for the new plaster, in parts even holes were pounded in the Carolingian paintings. Nevertheless, the adhesion between the older and the more recent plaster is poor in many places, which has caused parts of the newer paintings falling off the wall. Using the laser-optical measuring technique, it has been possible to identify large loose sections in many places, which can now be kept under close observation by conservators. This is demonstrated in Fig. 13 that shows the colour-coded result obtained from a wall in the south apse of the church. Again, loose sections are those areas that vibrated often and thus are indicated by red or yellow, the intact portions that could hardly be excited are coloured green to blue. Contours of the paintings are overlaid to indicate the location in the mural. At some points, it may be necessary to reinforce the connection between plaster and substrate. One alarm signal would be if these damaged sections were to become larger – possibly as a consequence of a minor earthquake that shook Graubünden in 2001. This will now be confirmed by comparing the earlier data with data from a repeat measurement. Here, an advantage over the traditional method of manual testing is its objectiveness which is important in such a comparison.

Additional information from the vibration data can be utilized for more detailed depth sounding. Thick layers respond at low, thin layers at higher frequencies. In a multi-layer plaster this allows speculations about the depth of the damaged interface. For an example, we present results of a study on a two-layer plaster coating at a historical wall of the Neues Museum, Berlin that was checked for the success of a remedy in which restorers had injected fixing cement through a certain number of small holes. Preliminary studies at a location where the upper layer was missing revealed that the lower layer responded mostly to frequencies below some 800 Hz. Thus, we grouped our results into two maps, one considering all data below, the other those above 800 Hz. In Fig. 14, hatching from upper left to lower right indicates regions where the lower interface (response to frequencies below 800 Hz) is considered loose; hatching from lower left to upper right indicates according regions for the upper interface (frequencies above 800 Hz). The width of the scene was 0.9 m; the bright dots give locations of cement injections. The results clearly show that the lower interface was not repaired as we find one large vibrating region in the low-frequency domain that includes the location of many of the fixation points. The response of the thinner upper layer (indicated by the many small regions in the high-frequency map) respects most of these points. Probably, the holes did not penetrate through

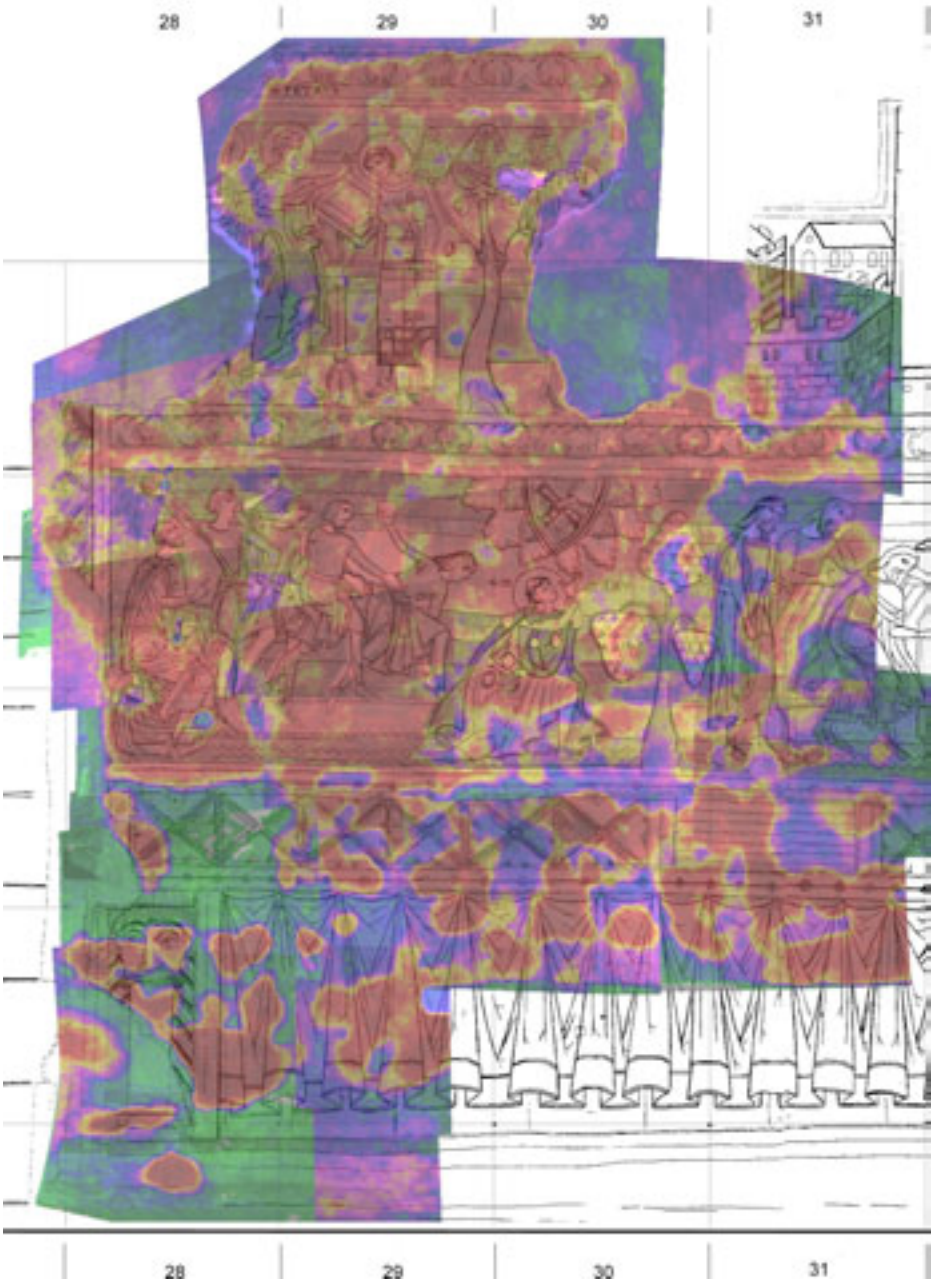


Figure 13. Map of loose areas in a Romanesque fresco of the south apse in St. John's Church of the World Heritage Site Müstair in Graubünden, Switzerland. Vibration ESPI results indicate intact regions by green/blue, loose areas by red/yellow. Height of the scene some 5 metres; the drawing describes the historical painting.

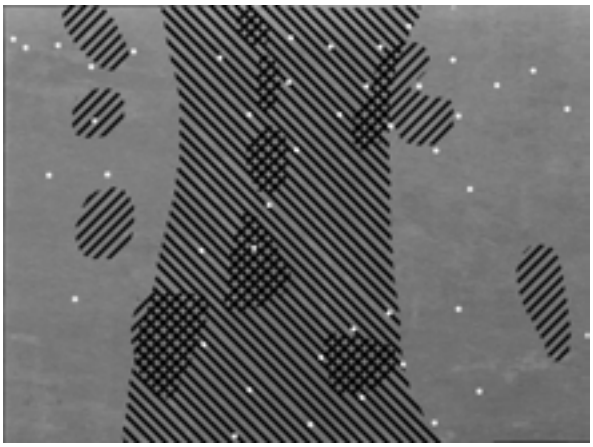


Figure 14. Localization of detachments in a two-layer plaster on a wall in Neues Museum, Berlin. Results of a series of ESPI measurements at equidistant frequencies. Hatching from upper left to lower right indicates regions where the lower interface (response to frequencies below 800 Hz) is considered loose; hatching from lower left to upper right indicates according regions for the upper interface (frequencies above 800 Hz). The bright dots give locations of cement injections.

the second layer and the cement connected only the two layers, but did not attach them to the wall.

7 Conclusions

Techniques of coherent optical metrology have proven well-suited for the investigation of the mechanical processes involved in the deterioration of artwork. They offer the necessary sensitivity to explore displacement fields and surface changes, and they are non-invasive which is of essential importance in delicate historical pieces. While the basic usefulness has been shown in various problems the metrology must now be developed for general acceptance as a tool in the everyday tasks of restorers and conservators.

Acknowledgements

This account is a summary of work that has been carried out mainly in the Applied Optics Group at the Institute of Physics in Oldenburg. The dedicated contributions of many members of the group throughout the years enter into the result which is acknowledged gratefully. Since the very beginning G. Gülker and H. Helmers have been involved in many of the studies. Special mention deserves the doctoral thesis work of J. Burke, T. Fricke-Begemann and H. Joost, who have contributed essentially to our progress in speckle metrology. Recent advances in LCSI are benefiting from cooperation with K. Gastinger at SINTEF, Trondheim, Norway. We also acknowledge the financial support from DFG, BMBF and DBU.

References

[1] D. Paoletti and G. S. Spagnolo, 'Interferometric methods for artwork diagnostics', in *Progress in Optics* (Elsevier, 1996), vol. XXXV, pp. 197–255.

[2] K. Hinsch and G. Gülker, 'Lasers in art conservation', Physics World **14**, 37 (2001).

[3] G. Gülker, H. Helmers, K. D. Hinsch, P. Meinlschmidt, and K. Wolff, 'Deformation mapping and surface inspection of historical monuments', Opt. Las. Eng. **24**, 183 (1996).

[4] H. J. Dainty, *Laser Speckle and Related Phenomena* (Springer, Berlin, 1975).

[5] M. Sjödahl, 'Digital speckle photography', in *Digital Speckle Pattern Interferometry and Related Techniques*, edited by P. K. Rastogi (Wiley, Chichester, 2001).

[6] K. D. Hinsch, G. Gülker, H. Hinrichs, and H. Joost, 'Artwork monitoring by digital image correlation', in *Lasers in the Conservation of Artworks*, edited by K. Dickmann, C. Fotakis, and J. F. Asmus (Springer, Berlin, 2004), LACONA V Proceedings.

[7] T. Fricke-Begemann, 'Three-dimensional deformation field measurement with digital speckle correlation', Appl. Opt. **42**, 6783 (2003).

[8] T. Fricke-Begemann and K. D. Hinsch, 'Measurement of random processes at rough surfaces with digital speckle correlation', J. Opt. Soc. Am. **A 21**, 252 (2004).

[9] R. Jones and C. Wykes, *Holographic and Speckle Interferometry* (Cambridge University Press, Cambridge, 1983).

[10] K. Creath, 'Phase-shifting speckle interferometry', Appl. Opt. **24**, 3053 (1985).

[11] J. Burke, *Application and optimization of the spatial phase shifting technique in digital speckle interferometry*, Dissertation, University of Oldenburg (2000).

[12] T. Bothe, J. Burke, and H. Helmers, 'Spatial phase shifting in electronic speckle pattern interferometry: minimization of phase reconstruction errors', Appl. Opt. **35**, 5310 (1997).

[13] M. Takeda, H. Ina, and S. Kobayashi, 'Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry', J. Opt. Soc. Am. **72**, 156 (1982).

[14] J. Burke, H. Helmers, C. Kunze, and V. Wilkens, 'Speckle intensity and phase gradients: influence on fringe quality in spatial phase shifting ESPI-systems.', Optics Comm. **152**, 144 (1998).

[15] J. Huntley and H. Saldner, 'Temporal phase-unwrapping algorithm for automated interferogram analysis', Appl. Opt. **32**, 3047 (1993).

[16] J. Burke and H. Helmers, 'Matched data storage in ESPI by combination of spatial phase shifting with temporal phase unwrapping', Opt. Las. Technol. **32**, 235 (2000).

[17] G. Gülker and K. D. Hinsch, 'Detection of surface microstructure changes by electronic speckle pattern interferometry', J. Opt. Las. Eng. **26**, 165 (1997).

[18] T. Fricke-Begemann and J. Burke, 'Speckle interferometry: three-dimensional deformation field measurement with a single interferogram', Appl. Opt. **40**, 5011 (2001).

[19] G. Gülker, K. D. Hinsch, and A. Kraft, 'Deformation monitoring on ancient terracotta warriors by microscopic TV-holography', Opt. Las. Eng. **36**, 501 (2001).

[20] K. Gastinger, G. Gülker, K. D. Hinsch, H. M. Pedersen, T. Stren, and S. Winther, 'Low coherence speckle interferometry (LCSI) for detection of interfacial instabilities in adhesive bonded joints', in *Proc. SPIE vol. 5532* (SPIE, Bellingham, 2004), p. 256.

[21] S. Ellingsrud and G. O. Rosvold, 'Analysis of a data-based TV-holography system used to measure small vibration amplitudes', J. Opt. Soc. Am. **A9**, 237 (1992).

[22] G. Gülker, K. D. Hinsch, and H. Joost, 'Large-scale investigation of plaster detachments in historical murals by acoustic stimulation and video-holographic detection', in *Proc. SPIE vol. 4402* (SPIE, Bellingham, 2001), p. 184.

Oscillations, Waves and Interactions, pp. 279–310
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-11-4

High-resolution Sagnac interferometry

K. U. Schreiber

Technische Universität München, Forschungseinrichtung Satellitengeodäsie, Fundamentalstation Wettzell, 93444 Bad Kötzting

Email: ulrich.schreiber@bv.tum.de

Abstract. Ring lasers are the most important sensors for the measurement of rotation when it comes to high stability and high sensor resolution. Their scale factor and hence their sensitivity increases with the area enclosed by two counter-propagating laser beams. Over the last decade a number of extremely large ring lasers were built, improving the sensitivity and stability of the measured rotation rate by several orders of magnitude over previous commercial developments. This progress has opened the door for entirely new applications of ring laser gyroscopes in the fields of geophysics, geodesy and seismology. Ring lasers for example are currently the only viable measurement technology, which is directly referenced to the instantaneous rotation axis of the Earth. This document reviews the research carried out by our international working group over the last decade and describes the current state of the large-scale ring laser technology.

1 Introduction

Highly sensitive rotation sensors have many applications. They reach from applications in robotics over navigation up to high-resolution measurements in seismology, geodesy and geophysics. The field of these applications is very broad and therefore a wide range of different sensor types and specifications exists to satisfy these demands.

In order to understand the importance of rotation sensors one should keep in mind that there are in total six degrees of freedom of movement, three for translations and three for rotations, respectively. While the measurements of translation are usually based on the determination of accelerations relative to an inertial test mass, rotations can be established either by mechanical gyroscopes or they can be measured absolutely by exploiting the Sagnac effect. Today fibre-optic gyros are the most prominent representatives for passive optical Sagnac interferometers, while ring laser gyroscopes represent the group of active Sagnac devices. They characterize the most sensitive and most stable class of gyroscopic devices.

2 History of Sagnac interferometers

In 1881 A. Michelson set up an L-shaped optical interferometer and showed subsequently that no ether could exist, provided the ether is assumed to be at rest and is

not moving along with the Earth. In order to investigate the possibility of a dragged ether, George Sagnac set up a different experiment. He generated a coherent beam of light, which he guided around a contour with a predetermined area of $0.086\,\mathrm{m}^2$. The entire apparatus was then rotated with a frequency of approximately $2\,\mathrm{Hz}$ [1]. With the help of a beam splitter and several mirrors he managed to generate two counter-propagating beams passing around the same optical path. He observed a shift in the interferogram of 0.07 ± 0.01 fringes and found that the measured shift was directly proportional to the rate of rotation. This observation, known as the '*Sagnac Effect*' today, however would require an ether at rest and was in contradiction with Michelson's findings. As a result of both experiments the ether theory was concluded.

A full description of the Sagnac effect is based on General Relativity [2], however in this case a classical interpretation yields the same result [3]. The observed phase difference is

$$\delta\phi = \frac{8\pi A}{\lambda c}\mathbf{n} \cdot \mathbf{\Omega}, \tag{1}$$

where A is the area circumscribed by the laser beams, λ the optical wavelength, c the velocity of light, $\boldsymbol{n}$ the normal vector upon A and $\mathbf{\Omega}$ the rate of rotation of the interferometer. Equation (1) relates the obtained phase difference to the rate of rotation of the entire apparatus and can be interpreted as the gyroscope equation [4].

Fibre-optic gyros (FOG) are modern representatives of this kind of optical gyroscopes. Because glass fibres with a length of several hundred meters are used, the scale factor can be made very large by winding the fibre to a coil and the sensitivity for rotational excitations is therefore much larger than for G. Sagnac's experiment. While the rotation rate of the Earth would have generated a fringe shift of as little as $\frac{1}{300}$ on his historic instrument, which was well outside the range of sensitivity, Earth rotation can be observed to about an accuracy of 10% even on relatively modest FOGs. Based on the experiment of G. Sagnac it was possible to estimate the required size of an instrument capable of resolving an angular velocity of $\approx 50\,\mu\mathrm{rad/s}$, which corresponds to the amount of Earth rotation experienced at mid-latitude.

In this context the famous experiment[1] of Michelson and Gale [5] in 1925 must be viewed. Figure 1 shows a design draft of this experiment. A beam path in an evacuated rectangular arrangement of pipes with a length of about 603 m by 334 m was used for that purpose. The incoming coherent light beam was split into two counter-propagating beams with the help of the beamsplitter A and then guided around the contour A, D, E and F by three more mirrors. The rotation rate of the Earth at the location of Clearing (Illinois) generated a shift of 0.23 fringes, measured with an uncertainty of no more than 0.005 fringes. This corresponds to a measurement error of only 2%. From a historical point of view it is very interesting to note, that this concept contained a substantial experimental challenge. Since the Earth rotation rate is absolutely constant at this level of sensor resolution, Michelson and Gale had to prove that the observed fringe shift was indeed a measurement quantity and not an artifact generated from multiple reflections in the interferometer

[1] Please note that the goal was to measure a very small, nearly constant, angular velocity. The experiment was not intended to proof Earth rotation as such.

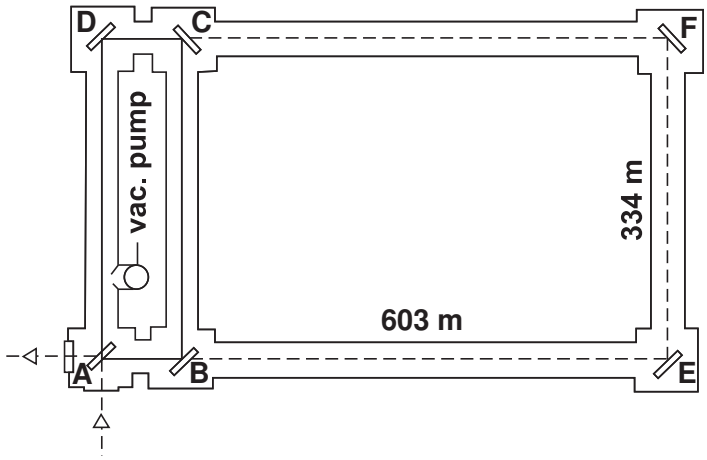


Figure 1. Sketch of the experiment of Michelson and Gale. The interferometer had a length of 603 m and a width of 334 m.

itself. Therefore they integrated a second much smaller interferometer corresponding to the contour A,B, C and D into their device, which created a negligible fringe shift. A comparison between the two different fringe pattern provided the required evidence.

3 Active Sagnac interferometers

In the days of coherent light sources it is possible to increase the sensor resolution of a Sagnac interferometer substantially. In particular the transition from a phase measurement to a frequency measurement promises a great improvement for the sensor sensitivity. By placing a laser gain medium inside, a closed light-path set up by three or four mirrors converts the apparatus into a laser with a ring cavity. Along with the lasing condition that the perimeter P corresponds to an interger number i of waves around the cavity, $P = i \cdot \lambda$, one obtains the ring laser equation

$$\delta f = \frac{4A}{\lambda P} \mathbf{n} \cdot \boldsymbol{\Omega} \,, \tag{2}$$

with δf the beat frequency, λ the wavelength of the optical beam, $\boldsymbol{n}$ the normal vector on the area, and $\boldsymbol{\Omega}$ the rate at which the entire apparatus rotates with respect to the universe. This beat frequency is usually called the *Sagnac frequency*. In the classical approach one can depict the co-rotating optical resonator as slightly larger compared to the anti-rotating resonator. Therefore both optical beams are shifted by the same small amount but in different directions away from the optical frequency which both beams would have when the apparatus were at rest.

Ring lasers have several advantages. First of all there are no moving parts in the design of this sensor. Secondly, they are sensitive to a very large range of rotation rates, covering more than six orders of magnitude with a linear system response.

These properties made them very welcome for applications in navigation. Compact designs, made monolithically from blocks of Zerodur, were used to make ring lasers of an area around $0.01\,\mathrm{m}^2$. They are widely used in aircrafts [6]. The sensitivity of a ring laser gyroscope mostly depends on the scaling factor, i. e. the ratio of the area and the perimeter. Today, 80 years after the outstanding experiment of Michelson and Gale, a number of gyroscopes have been built, which exceed the performance of the historic instrument by approximately two million, while the corresponding sensors are a lot smaller at the same time. This does not only allow one to measure the rotation rate of the Earth $\boldsymbol{\Omega}_\mathrm{E}$, but it also readily shows small variations of this quantity. Apart from a response of the Earth to external forces, these fluctuations are resulting from a momentum exchange between the atmosphere, the hydrosphere and the lithosphere. Therefore perturbations of Earth rotation can be used as an indirect indicator for the monitoring of global phenomena, such as variations in global ocean circulation.

4 Ring laser design

Atmosphere and hydrosphere make up a very small portion of the entire Earth mass only. Therefore an extremely high resolution for any type of gyroscope is required to access the information contained in these geophysical signals at a level of well below 10^{-7}. Apart from the actual sensitivity of the gyro also a remarkable sensor stability is required. Some known periodic signals, such as the Chandler wobble, have periods of around 432 days. Other such small-scale signals of interest are expected to be aperiodic and one would wish to distinguish them clearly from sensor drifts. In order to provide the required sensor stability it is desirable to make them mechanically as rigid as possible. Furthermore the shape must not be affected by ambient temperature and atmospheric pressure changes.

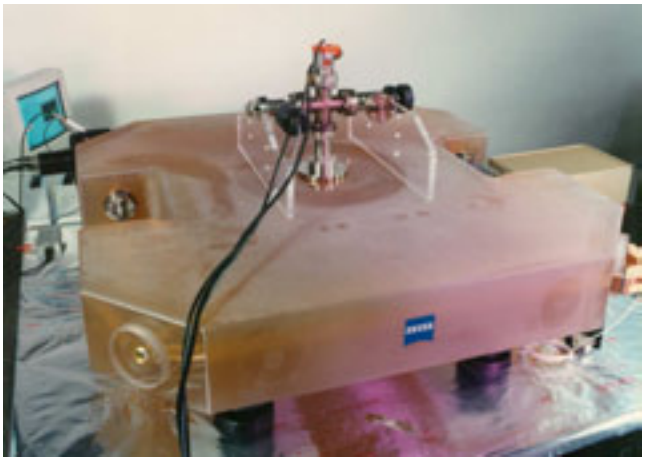


Figure 2. The C-II ring laser at the time of construction at Carl Zeiss (Oberkochen) at the end of the year 1996. Photograph courtesy of Carl Zeiss GmbH, Oberkochen.

Figure 2 shows an example of such a construction. C-II is a second generation Helium-Neon ring laser [7]. The body of the gyro is made from a slab of Zerodur, 18 cm thick. All four corners are bevelled and polished so that discs of ULE with optically contacted super mirrors can be wrung onto the ring laser body in order to generate a pre-aligned closed light path. The beam path itself is drilled into the neutral plane of the Zerodur slab, parallel to the sides, so that an area of $1\,\mathrm{m}^2$ is circumscribed by the laser beams. On one side, half way between the mirrors, there is a cut-out in the ring body. A small adjustable capillary with a diameter of 4 mm is placed in this gap. Two electrical loops around it act as an rf-antenna. They are used to excite a gas discharge for laser excitation. A similar cut-out on the opposite site, closed up with a 12 mm wide pyrex tube, is also integrated into the design for future experimental purposes. There are also two diagonal holes drilled through the entire ring laser structure. A UHV-valve located in the center above the Zerodur block seals the cavity off from the environment. This valve can be connected to a turbo pump in order to evacuate the ring cavity. Subsequently to the pumping, a mixture of Helium and Neon can be applied to fill the entire inner part of the construction with a few mbar of total gas pressure. The design of these ring lasers was chosen such, that the quality factor of the cavity, Q, was made as large as possible. There are no Brewster windows or mode selecting devices contained in the entire setup [8]. Together with the very low loss from the mirrors ($\approx$ 20 ppm per mirror) a Q on the order of 10^{12} and higher is achieved for all the rings. The cavity quality factor Q, defined as $Q = \omega\tau$, was established from decay-time measurements. High values for Q result in a narrow linewidth of the laser and, equally important, in a much reduced systematic offset of the Sagnac frequency from its true value due to the lock-in effect.

Essentially all large ring lasers built by the German-New Zealand collaboration follow the same design principle. They are He-Ne gas lasers with rf excitation, optimized for low loss operation. Since all viable active Sagnac interferometers have to operate on a single longitudinal mode per sense of propagation, all the devices are operated near the laser threshold. A free spectral range (FSR $= c/P$) between 2.4 MHz and 75 MHz would otherwise allow many different longitudinal modes to oscillate. The cross-section of the capillary used for laser excitation together with the radius of curvature of the mirrors determines at which transversal mode laser oscillation comes on. Preferably all rings operate on TEM-0,0.

Figure 3 shows the G ring laser during an upgrade of the vacuum tubing in 2006. The dimensions of G are 4 m by 4 m and up to today this is the most stable and most sensitive sensor. While the smaller rings C-II and G are of monolithic construction, UG-1 ($367.5\,\mathrm{m}^2$) and UG-2 ($833\,\mathrm{m}^2$) are much too large for that. Therefore they have a heterolithic design. The corners of these ring lasers, the laser gain section and the connecting vacuum tubes are resting on small pedestals around the perimeter of the Cashmere Cavern in the northern slope of the Banks Peninsula south of Christchurch in New Zealand. Figure 4 gives an impression about the construction of these very large rings.

Since ring lasers are also very suitable sensors for the monitoring of rotations induced by earthquakes, a simplified version of an active Sagnac interferometer was constructed in order to provide a relatively cheap but still sensitive device to the seismological community. As there is no real need for long-term stability, this GEOsensor

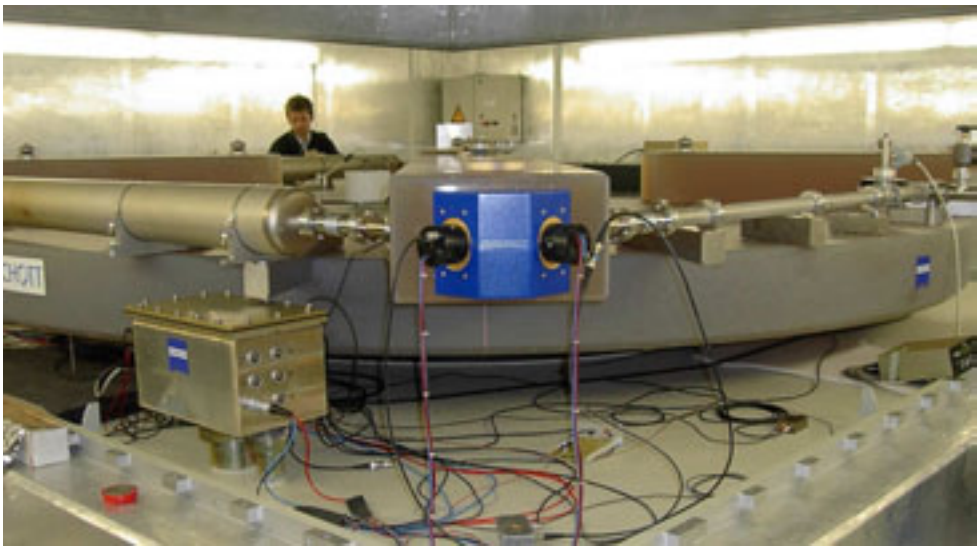


Figure 3. The G ring laser during an upgrade of the vacuum system at the Geodetic Observatory Wettzell in 2006.

Figure 4. Illustration of the heterogeneously constructed UG2 gyroscope. Many small concrete foundations on the floor of the Cashmere cave support the ring laser structure.

does also not require Zerodur for the interferometer body. The application of ring lasers to seismology is covered in more detail in Sect. 7.

In order to summarize the basic performance of all important large rings, Table 1 lists all relevant quantities from these instruments. Apart from area and perimeter, the key figures for the scaling factor, the table also lists the obtained Sagnac splitting of the optical frequency as well as the best short-term sensor resolution obtained from these devices. It should be noted that all rings are orientated horizontally on

Ring Laser	Area, m^2	Perimeter, m	f_{Sagnac}, Hz	$\Delta\Omega/\Omega_{\mathrm{E}}$
C-II	1	4	79.4	$1 \cdot 10^{-7}$
GEOsensor	2.56	6.4	102.6	$1 \cdot 10^{-7}$
G0	12.25	14	288.6	$4 \cdot 10^{-7}$
G	16	16	348.6	$1 \cdot 10^{-8}$
UG1	366.83	76.93	1512.8	$3 \cdot 10^{-8}$
UG2	834.34	121.435	2177.1	$5 \cdot 10^{-8}$

Table 1. Summary of physical properties of a number of large ring lasers.

the Earth with the exception of G0, which is located vertically along an east/west wall. For this table only one parameter, namely the theoretical sensitivity, has been regarded. This does not necessarily mean that the performance is readily obtained at all times. The enhanced scaling factor comes on the expense of mechanical stability and also a faster degradation of the laser gain medium over time. In order to draw conclusions from ring laser measurements with respect to global geophysical signals it is important to rigidly connect the interferometer to the Earth crust. Naturally it is easier to do that with larger constructions. However, a heterolithic structure inside an artifical cave suffers substantially from thermo-elastic deformations and atmospheric pressure variations. The following section looks at the various error mechanisms of ring lasers in more detail.

5 Sensor resolution

According to Refs. [4, 15] and also others the sensitivity limit of a ring laser gyroscope from the irreducible quantum noise for a rotation measurement is given by

$$\delta\Omega = \frac{cP}{4AQ}\sqrt{\frac{hf}{P_x t}}\,, \tag{3}$$

where P is the perimeter, A the area encircled by the light beams of the gyro, $Q = \omega\tau$ the quality factor of the ring cavity, h is Planck's constant, P_x the beam power loss corresponding to the photon flux on the photodetector and t the integration time. For the large ring laser G in Germany $P = 16\,\mathrm{m}$ and $A = 16\,\mathrm{m}^2$. The ring-down time was first measured to be $\tau = 1\,\mathrm{ms}$ in 2001. Over the years it reduced to a value of $\tau = 500\,\mu\mathrm{s}$ in 2007 due to gradual mirror degradation. Figure 5 shows the current quantum limit for G as a function of the respective integration time. Currently G reaches a sensitivity of $10^{-12}\,\mathrm{rad/s}$ at an integration time of approximately 1000 seconds, which is believed to be a world record.

All these ring lasers are very large compared to an aircraft gyro. The optical path length varies between several meters and 121.44 m. With total cavity losses at the level of 108 parts per million, this translates into a very narrow linewidth. The theortically expected Schawlow-Townes linewidth for G is

$$\Delta\nu_L = \frac{N_2}{N_2 - N_1}\frac{2\pi f_0 \Delta\nu_c}{P_L}\,, \tag{4}$$

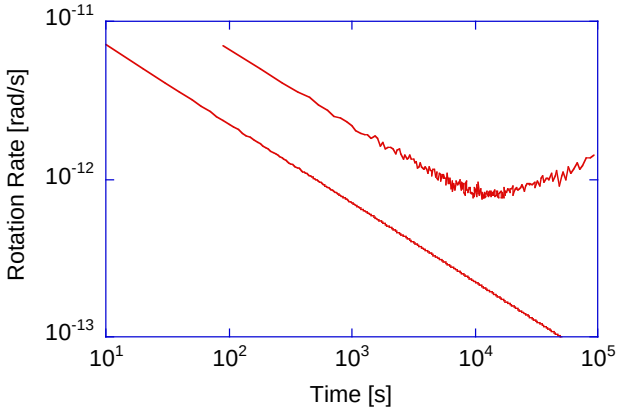


Figure 5. Computed quantum noise limit and measured sensor resolution for G. The sensor almost reaches the shot noise limit.

with a width of the cavity resonance of

$$\Delta_{\nu_c} = \Delta_{\mathrm{FSR}} \frac{1-R}{\pi\sqrt{R}} \,. \tag{5}$$

Using $\Delta_{\mathrm{FSR}} = c/L = 18.75\,\mathrm{MHz}$ and $R = 0.999892$ one obtains $\Delta\nu_c \approx 645\,\mathrm{Hz}$ and hence $\Delta\nu_L \approx 275\,\mu\mathrm{Hz}$. Since a He-Ne laser is a 4-level system one may expect to set $N_1 \approx 0$ which assumes full inversion. For an industry type He-Ne laser with a capillary diameter of 1-2 mm and a moderate excitation current this assumption is certainly true. However, our lasers use wider capillaries with 4 mm (C-II), 5 mm (G) and 6 mm (UG2) diameter. This slows down the wall-collision induced de-excitation of the Neon atoms from the $1s$ state. As a result the laser gain reduces substantially and, by the process of electron-Neon collision, pumping leads to a an increase of N_1, a reduction of the inversion [3, 16], and consequently to an increased linewidth. In the absence of any variations of the scaling factor according to Eq. (3) one obtains a constant lower limit for the gyro resolution, which then is only depending on fluctuations of the laser beam power in the cavity. A typical value for the stability of the beam power due to mode competition of 0.01% over a time period of about 1 hour has been observed from intensity measurements in G.

In order to investigate the mid-term stability of large ring lasers, Fig. 6 shows an Allan deviation plot of most of the lasers specified in Table 1. One can see that the monolithic constructions are more stable than the heterolithic structures. This is not unexpected. Compared to UG1, C-II experiences much more perturbations from backscattering and comes with a much smaller scale factor. The performance of UG2 on the other side falls off despite the enhanced scale factor. This may be due to the rectangular sensor layout or comes with the increased arm length of the instrument. G0 as the only vertically mounted ring laser does not have the mechanical stability to perform well. Actually this prototype ring laser construction was only intended to test the feasibility of heterolithic ring laser designs. The GEOsensor is designed

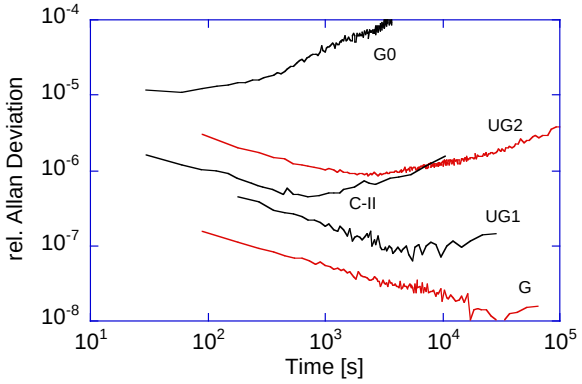


Figure 6. Relative Allan deviation of most of the large ring lasers of Table 1. One can see that UG2 is less stable than UG1. This may be due to the rectangular shape or the scale factor was pushed beyond the practical limit for this type of construction.

specifically for studies in the field of rotational seismology, where long-term stability is not a design criterion.

6 Error contributions

Several mechanisms in a real world ring laser cause a departure of the actually measured Sagnac frequency from the theoretical value. In a much generalized form one can write

$$\Delta f = K_R(1 + K_A)\mathbf{n} \cdot \mathbf{\Omega} + \Delta f_0 + \Delta f_{bs}\,, \tag{6}$$

where $K_R = 4A/\lambda L$ is the geometrical scaling factor of an empty ring laser cavity. The quantity K_A accounts for the additional contributions due to the presence of an amplifying laser medium, while Δf_0 allows for mode pulling and pushing because of dispersion, and Δf_{bs} takes the coupling of the two laser beams in the presence of backscatter into account [9]. These latter two effects are well established in the ring laser literature (e. g. [10, 11]) and are usually both very small and almost constant for the ring lasers discussed here. Ring laser applications in geodesy and geophysics require ultimately stable and highly sensitive sensors with a demand for a relative sensor resolution of $\Delta\Omega/\Omega_{\mathrm{E}} < 10^{-8}$. Therefore it is important to understand the nature and variability of these error contributions.

6.1 Non-reciprocal effects in the laser cavity

For an ideal ring laser both of the two counter-propagating laser beams would be identical with respect to beam size and intensity. However in practise there is a noticeable difference in the beam intensities for all of the large ring lasers. Differences in the intensities of more than 20% have been found for C-II, and even G shows a beam power difference of more than 10%. The reason for that is not fully understood

and it is currently believed that the respective mirror coatings exhibit some minute non-isotropy, causing the cavity Q to be different for the two senses of propagation.

For a model accounting for dispersive frequency detuning and hole burning [12] the bias due to the corresponding null shift offset becomes

$$\Delta f_0 = \frac{c}{2P} \cdot \left(\frac{\xi}{\eta} L(\xi) \frac{Z_i(\xi)}{Z_i(0)} G \right) \cdot \Delta I \,, \tag{7}$$

with ξ the cavity detuning from line center, $Z(\xi)$ the imaginary part of the dispersion function, G the gain, and ΔI the observed difference in intensity. Equation (7) also shows that it is very important to keep the gain constant, which is very difficult for a gas laser in the presence of gain medium degradation through outgassing inside the cavity.

6.2 Backscatter coupling

In the presence of strong backscattering, light of one sense of propagation is coupled into the beam travelling into the opposite direction. According to Ref. [9] the beat frequency disappears if the experienced rotation rate falls below a threshold value of

$$\omega_L = \frac{c\lambda^2 \sqrt{r_s}}{32\pi A d} \,. \tag{8}$$

A is the area enclosed by the cavity, d the diameter of the beam, and r_s the amount of backscatter contribution of the mirror. In the most general case the shifted beat frequency then becomes

$$\Delta f = \frac{4A}{\lambda P} \sqrt{\omega^2 - \omega_L^2} \,, \tag{9}$$

where ω is the experienced rate of rotation of the gyro. For ring lasers of the size of C-II or smaller, ω_L can be of a significant amount, however far from locking up with the Earth rate as the only source of rotation, while neither G nor any of the UG ring lasers ever showed much evidence for the presence of backscatter at all. For this situation the contribution to backscatter is best expressed as

$$\Delta f_{bs} = \frac{c}{2P} \left(\rho_2 \sin(\psi + \epsilon_2) + \rho_1 \sin(\psi + \epsilon_1) \right) \,, \tag{10}$$

where ϵ_1 and ϵ_2 are the respective backscatter phase angles and ρ_1 and ρ_2 the corresponding backscatter amplitudes. Following Ref. [12] one can write

$$\rho_1 = \frac{r_s \lambda}{4d} \sqrt{\frac{I_1}{I_2}} \,, \quad \rho_2 = \frac{r_s \lambda}{4d} \sqrt{\frac{I_2}{I_1}} \,, \tag{11}$$

with I_1, I_2 the respective intensities of the two beams. As one can see from Eqs. (10) and (11) together, the effect of backscattering reduces with growing size of the cavity and lower scattering amplitude. In particular, the reduction of backscattering through a limited acceptance angle of the solid angle where all the light was scattered into, seems to be a very effective process.

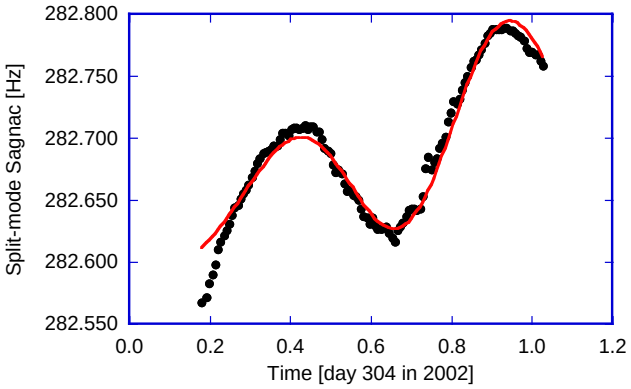


Figure 7. Earth strain changes the perimeter in UG1. The black dots show the variation in the FSR, while the line represents the expected variations from a global strain model. An additional linear term accounts for the increase of temperature in the Cashmere Cavern over the course of these measurements.

6.3 The effect of Earth strain on large cavities

Large ring laser cavities such as UG1 and UG2 are subject to Earth strain effects caused mainly by the gravitational attraction of the moon. The deformation in response to the lunar gravitational pull stretches the cavity with an amplitude of about 20 nanostrain, which adds up to a total of approximately 1.5 μm for UG1, thus changing the geometrical part of the scale factor $4A/\lambda P$ by a small amount. In the absence of any shear forces these strain effects will change both the perimeter and the area of the ring laser at the same time. Since the perimeter has to contain an integer number of waves ($P = I\lambda$) in order to maintain lasing, the scale factor can be reduced to $(I/4)$ for a square ring, where I is the longitudinal mode index. This means that in this case all the changes in the area and perimeter are compensated by a corresponding shift in the optical frequency of the laser, provided that the shift is less than one free spectral range and the laser mode does not change. While this self-compensation is strictly speaking only valid for a true square ring, one can find that it also applies to near square rings.

Figure 7 gives an example from UG1. For the purpose of this measurement UG1 was operated on 2 neighbouring longitudinal modes per sense of propagation. The beat note between these two modes $\Delta f = c/P \cong 3.9\,\mathrm{MHz}$ corresponds to the free spectral range and is a direct measure of the effective length of the cavity. The beat frequency was down-converted to an audio signal of about 288 Hz with the help of a GPS controlled reference oscillator and continuously recorded with an A/D-converter and a computer. While the black dots in Fig. 7 represent the variation of the length of the cavity expressed as a shift of the measured FSR, the red line gives the corresponding FSR shift computed from an Earth strain model. In order to account for an upward drift of the ambient temperature in the Cashmere Cavern over the course of the measurements an additional linear drift term was added. A

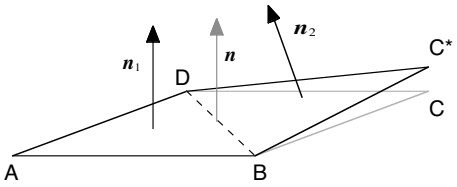


Figure 8. Determination of the orientation of a non triangular ring laser. While $\boldsymbol{n}$ represents the orientation of the square ABCD, a non-planar ring has to be subdivided into triangles, which then have to be summed up.

good agreement between the model and the FSR measurements is found, while the scale factor variations coming from the strain effects are not visible in the time series of the Sagnac frequency.

6.4 Geometric scale factor correction

The two monolithic smaller ring lasers and their monuments are geometrically very stable and beamwalk effects have not been observed so far. In contrast the very large ring lasers UG1 and UG2 are subject to deformations of the Cashmere Cavern as a result of thermoelastic strain and atmospheric pressure variations. Small mirror tilts in the laser beam steering cause changes in area and perimeter. This results in a drift of the geometrical scaling factor and the normal vector of the respective ring laser. Other systematics come from the fact that the gyroscopes are He-Ne gas lasers and therefore they suffer from a continuous degradation of the laser gas purity caused by outgassing from the cavity enclosure. As the laser gain reduces with time a substantial drift of the measured Sagnac frequency develops. The obtained beat frequency from a ring laser gyro is proportional to the scaling factor, the rotational velocity and the orientation of the area normal vector and the vector of rotation as shown in Eq. (2). The normal vector on the plane of the laser beams is well defined for a triangular ring only.

Since most of the very large ring lasers existing to date have a square or rectangular shape, one needs to modify the definition of orientation for these instruments. Figure 8 outlines the procedure. On the assumption of a square ring laser completely planar along the corners ABCD, one finds the normal vector $\boldsymbol{n}$ representing the entire area. If however one corner (C^*) is slightly tilted out of plane, the effective area may be obtained by subdividing the full area into triangles and projecting the normal vector $\boldsymbol{n_i}$ of each triangle onto the vector of rotation and summing them up.

Nearly all the large ring lasers mentioned in Table 1 are orientated horizontally on the Earth. Because Earth rotation is the most dominant measurement signal, they show a strong latitude dependence of $\Omega_{\mathrm{eff}} = \sin(\phi + \delta_N)$, with ϕ corresponding to the latitude of the instrumental site and δ_N representing a tilt towards North. East-west tilts are nearly negligible, since the cosine of an angle representing a small tilt towards East $\delta_{\mathrm{E}} \approx 0$ is so close to 1 that it can be neglected, except for strong seismic motions.

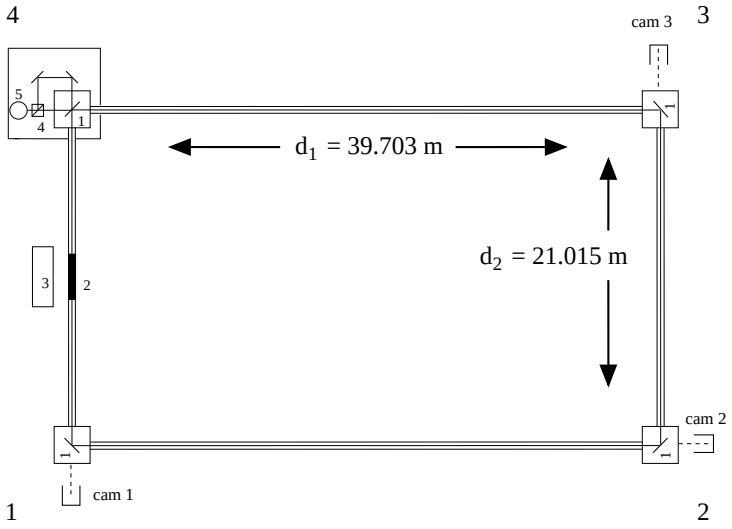


Figure 9. Basic layout of the UG2 ring laser. The laser beams are steered around the cavity by mirrors with precise mounting on the corner monuments.

UG2 has a rectangular layout spanning an area of 39.703 m by 21.015 m. The basic design is shown in Fig. 9. Because of the long beam trajectories of 39.703 m and 21.015 m of the laser, small mirror tilts in the range of a few seconds of arc are causing already a noticable beam displacement on the next mirror. Since this corresponds to a change in the geometric scale factor, the beamwalk was monitored and the instantaneous area and ring laser orientation (relative to the ring laser hardware structure) was computed. By placing a CCD camera behind the mirrors at the locations 1–3 indicated in Fig. 9 and recording the light leakage of the laser beam through the mirrors, beamwalks on the order of a micrometer in displacement could be monitored. Measurements were conducted by averaging over 4000 individual images taken one after the other, with the maximum supported exposure time of the cameras of 8 ms. Figure 10 shows a sample measurement sequence of the movement of the beam spot center postion with time. In this particular dataset the excursion of the horizontal position of the laser beam is two times larger compared to the vertical movement. On other occasions both movements were of the same proportion.

Since the determination of scale factor variations are required with a relative precision of 10^{-9} one can expect a substantial improvement from the geometrical scale factor correction. After the alignment of the UG2 ring laser to optimize it for minimum losses, and after the laser was refilled with a clean supply of Helium and Neon, a measurement sequence of approximately two weeks was started on Dec. 30, 2005 and lasted until Jan. 15, 2006. The Sagnac frequency was recorded with an integration time of 30 minutes. Figure 11 shows the measurement sequence of raw data as recorded on the logging system. From the experimental setup, one would expect

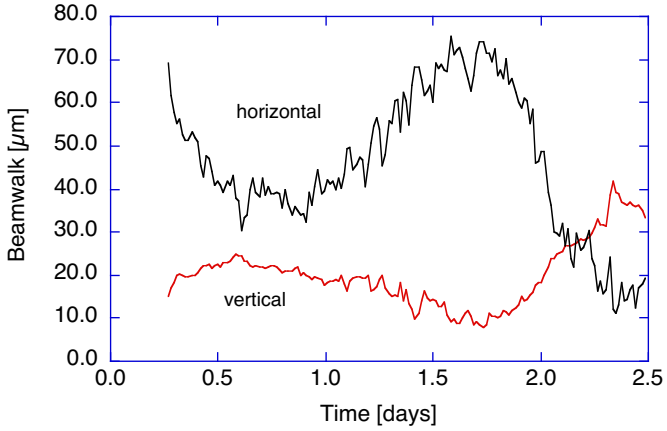


Figure 10. Example of the measured beam wander as obtained at corner 1. Within a factor of two vertical and horizontal excursions are roughly of the same magnitude.

Earth rotation to produce a constant Sagnac frequency, with a variation of 1 mHz or less. However the measurements show an overall downward trend and superimposed systematic excursions of considerable amplitude, where some occur rather sharply, while others are following a much smoother course. Some of these departures from the value of Earth rotation are due to variations of the scale factor as a function of time while others are caused by internal processes in the laser cavity.

From the instantaneous beam postions recorded by the cameras the geometrical variation in area and perimeter can be computed via an ABCD-Matrix approach. Figure 12 shows the result. One can see that the area is changing at the parts per million level, certainly a considerable effect. The simultaneously displayed time

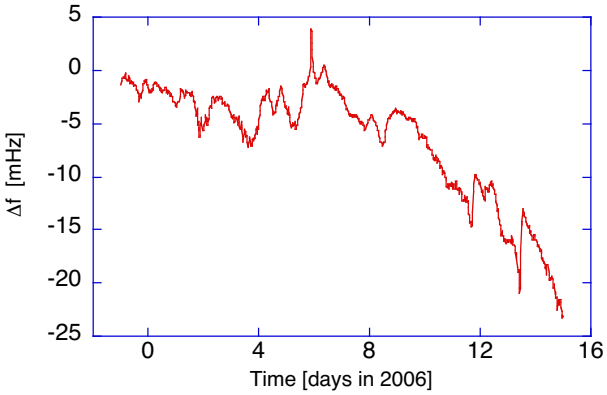


Figure 11. Time series of the drift of the raw measurements of Earth rotation obtained from the UG2 gyroscope.

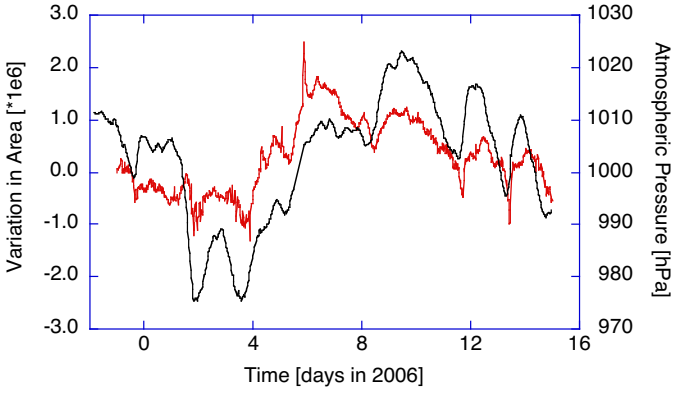


Figure 12. Time series of the variations in effective ring laser area superimposed with the atmospheric pressure at that time.

series of the atmospheric pressure as measured inside the cavern shows significant correlation with these changes in area. From that one may conclude that pressure induced deformations of the cave are causing small tilts at the mirror mounts, which in turn cause beamwalk on the next mirror. A similar result is obtained for the instantaneous orientation of the UG2 ring laser as shown in Fig. 13. The contributions from ring laser reorientation are smaller by a factor of about two compared to the variations in area. Nevertheless the corrections to the orientation vector are also in the parts per million regime and can not be neglected. Again there is some correlation with the atmospheric pressure evident as well as a linear overall trend. As opposed to orientation and area, the computed perimeter changes are approximately six orders of magnitude smaller as shown in Fig. 14. This is compatible with the general observation that longitudinal mode index changes are infrequent.

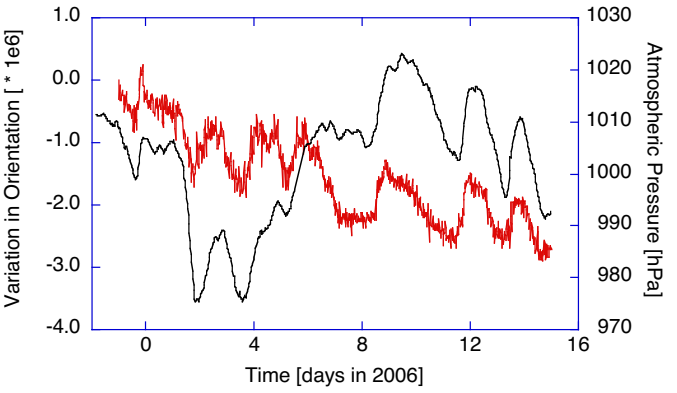


Figure 13. Time series of the variations in effective ring laser orientation superimposed with the atmospheric pressure.

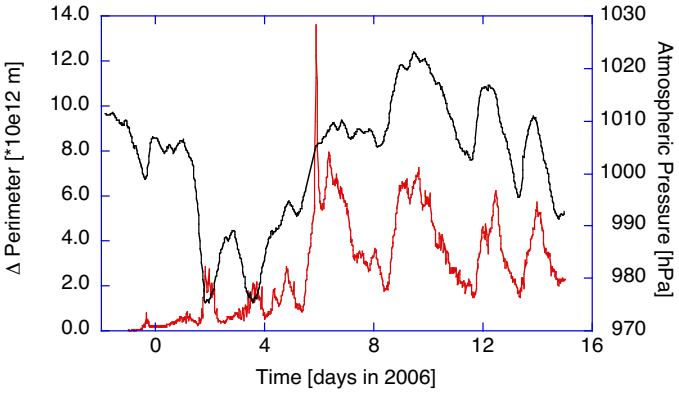


Figure 14. Time series of the variations of the UG2 ring laser perimeter superimposed with the atmospheric pressure.

Finally the complete instantaneous geometrical scale factor was computed as a function of time for the entire period of the measurements. One can see from Fig. 15 that the changes amount to several parts per million for UG2. This certainly can not be ignored for the analysis of ring laser measurements of Earth rotation. Similar measurements on the monolithically constructed G ring laser suggest that the scale factor variations from beamwalk are at least two orders of magnitude smaller and outside the range of resolution of the available cameras. When the computed corrections are applied to the UG2 raw data (Fig. 11) the excursions reduce substantially as shown in Fig. 16. As a result one can see that the amplitude of the departures from the expected Earth rotation rate have reduced substantially but they have not completely disappeared. Furthermore there is a strong systematic downwards drift in the Sagnac frequency contained in the residuals of the corrected measurements. The

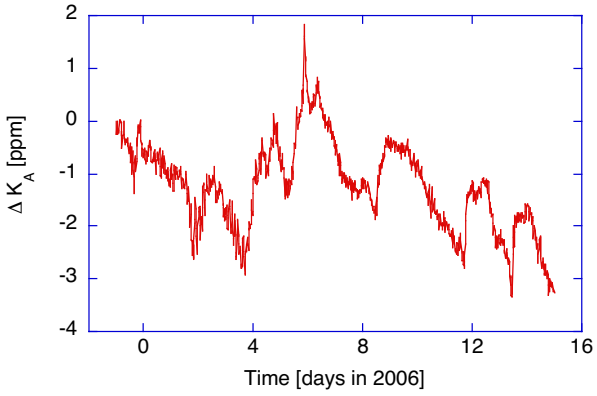


Figure 15. Time dependence of the variation of the ring laser scale factor over the period of the measurements.

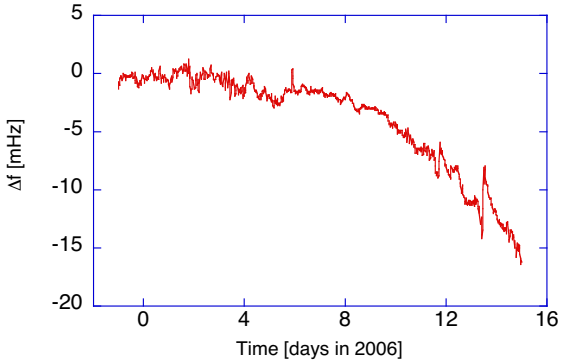


Figure 16. Time dependence of the variation of the ring laser scale factor over the period of the measurements.

systematic trend suggests that this unwanted signal may be introduced by the decay of the laser gain medium due to changes in dispersion as addressed in Section 6.1.

6.5 Scale factor corrections from varying laser gain

Apart from a major disruption around day 13 there is only a systematic downward drift roughly of the form $-e^t$ left in the data set. The existence of such a distinct systematic feature suggests the presence of an independent bias mechanism in a ring laser cavity. The residual downward trend is most likely due to a contamination of the laser gas with hydrogen, oxygen, nitrogen and water vapour via outgassing from the cavity enclosure and gives rise to an additional time dependent loss factor in the gain medium. This effect has been observed in a different context for linear He-Ne lasers, see Refs. [13, 14]. UG2 with a stainless steel tube enclosing the laser gas along the entire perimeter of 121.435 m experiences a substantial amount of outgassing. On two occasions measurements of the total gas pressure inside the cavity were made, each over a period of 56 days. An overall increase of 0.022 mbar and 0.020 mbar of hydrogen, amounting to $4 \cdot 10^{-4}$ mbar H_2 per day was observed, taking the UG2 ring laser well into the regime where additional losses from absorption effects by hydrogen become visible. Because of the need of adjusting large ring lasers to single longitudinal mode operation near laser threshold, a feedback circuit is used to stabilize the gain medium to constant circulating beam power. Growing losses in the laser cavity therefore will raise the loop gain accordingly. Following the approach of Ref. [10] the quantity K_A describes the contribution of the active medium to the scale factor of a large ring laser gyro as

$$\frac{\Delta K_A}{K_A} = \left(\frac{\Delta K}{K}\right)_N - \frac{aG}{1 + xP_o} + NL(\Omega)\,. \tag{12}$$

In this equation $\Delta K_A/K_A$ corresponds to the scale factor correction due to the active laser medium, $(\Delta K/K)_N$ is the constant part of the scale factor correction, the second term on the right-hand side allows for laser gain related contributions

and the last term accounts for nonlinear contributions such as backscatter-related coupling, which are neglected in the following discussion.

Usually, the gain factor G is considered constant with respect to time in ring laser theory. However, for the reasons outlined above, one has to account for the progressive compensation of gas impurity related losses by setting for example

$$G = G_0 e^{\alpha t} . \tag{13}$$

This choice of G is arbitrary and motivated from the behaviour of the loss as shown in Ref. [13]. The beam output power of Eq. (12) in the required form for large ring lasers with an 1:1 isotope mixture of ^{20}Ne and ^{22}Ne becomes

$$P_o = 2 I_s A_b T \left(\frac{G}{\mu} \cdot \frac{\kappa_1 Z_i(\xi_1) + \kappa_2 Z_i(\xi_2)}{Z_i(0)} - 1 \right) , \tag{14}$$

with I_s the saturation intensity, A_b the beam cross section and T the transmission of the laser mirrors. Z_i is the imaginary part of the plasma dispersion function with lasing at a frequency detuning of ξ_n with respect to the corresponding line centers of the two Neon isotopes, each having a partial pressure of κ_1 and κ_2, respectively. The most important part in this equation is the factor G/μ, which represents the gain–loss ratio. This factor is approximately constant over the time of the measurements because of the feedback loop operation. Since there is no drift in the optical frequency involved, the contribution of the plasma dispersion function also remains constant. Therefore, the denominator in Eq. (12) can be approximated by $1 + x P_o \cong 1$ so that this equation reduces to

$$\frac{\Delta K_A}{K_A} = \left(\frac{\Delta K}{K} \right)_N - a G_0 e^{\alpha t} . \tag{15}$$

Applied to the dataset of Fig. 16 one obtains a corrected dataset as shown in Fig. 17 after a nonlinear fitting procedure is performed. The result from Fig. 17 outlines the best mid-term performance obtained from UG2 so far.

6.6 Ring laser orientation

The Sagnac frequency in Eq. (2) contains contributions from three distinctly different mechanisms. Most obviously it relates the experienced rotation rate with the frequency difference observed between the two senses of propagation in the ring laser. Variations of the scale factor modify the proportionality between rotation rate and observed Sagnac frequency. The inner product between the normal vector on the ring laser plane and the vector of rotation finally determines how much of the rotation rate is projected onto the ring laser. For a large instrument rigidly attached to the ground and monitoring the Earth rotation rate this means, that any changes in orientation between the ring laser and the instantaneous Earth rotation vector show up in the Sagnac frequency.

Solid Earth tides and diurnal polar motion cause such orientation changes. A detailed description is given in Refs. [21, 22]. The direction of the Earth's rotation axis varies with respect to both Earth- and space-fixed reference systems. The principal

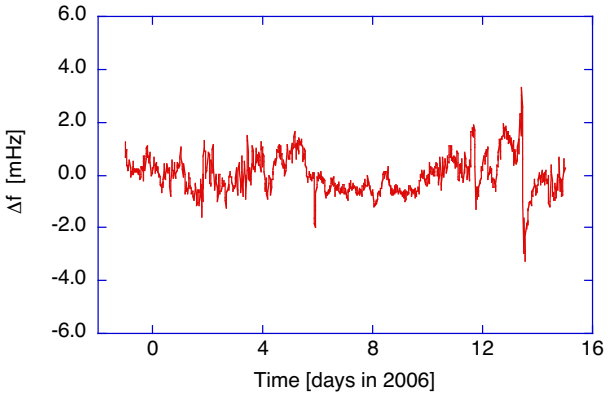


Figure 17. Time series of the UG2 Earth rotation measurements after the correction for geometrical scale factor variations as well as variations in gain from outgassing in the cavity.

component with respect to the Earth-fixed frame is the well-known Chandler wobble, with an amplitude of 4–6 m at the poles and a period of about 432 days. This is a free mode of the earth, i. e., it would still be present in the absence of the external gravitational forces. It is believed that the Chandler wobble would decay due to dissipative effects in the Earth's interior, were it not continually excited by seismic activity and by seasonal variations to atmosphere and ocean loadings. The Chandler wobble is overlaid by daily variations whose amplitudes are an order of magnitude smaller, some 40–60 cm at the Earth's surface [17, 18]. These wobbles arise from external torques due to the gravitational attraction of the Moon and Sun. Since the Earth is an oblate spheroid with an equatorial bulge which is inclined to the plane of the ecliptic, the net gravitational torques of the Moon and Sun on different parts of the Earth's surface do not exactly cancel out as it would be the case if the Earth were a perfect sphere. In an Earth-fixed reference system, such as that of a ring laser fixed to the Earth's surface, the forced retrograde diurnal polar motion is best viewed as a principal mode – the so-called "tilt-over mode" (K_1) – with the period of exactly one sidereal day (23.93447 hours), whose amplitude is modified as the angles and distances between the Earth, Moon and Sun vary over the course of their orbits.

The complete spectrum of nutation modes can be understood as the beat frequencies of the tilt-over mode with frequencies corresponding to relevant orbital parameters: half a tropical month, half a tropical year, the frequency of perigee etc. The beat periods are clustered around one sidereal day. The O_1 and P_1 modes, with beat periods of 25.81934 and 24.06589 hours, have the largest amplitudes after the K_1 mode, and arise from the change in angle between the Earth's equatorial bulge and the Moon and Sun respectively (see Ref. [19] for basic theoretical details.). Figure 18 shows an example of such a measurement series from G. The data in the diagram was reduced by subtracting the mean value of the Sagnac frequency from the dataset in order to give a better representation of the amplitude of the diurnal polar motion signal. The theoretically expected polar motion signal, visible as a north/south tilt of the Earth rotation vector, was computed from a model [20] and is also shown in the

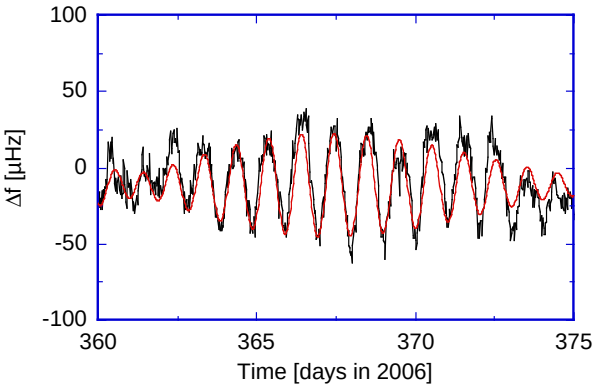


Figure 18. Time series of the variations of the Earth rotation measuremets of G after the mean value of the rotation rate has been subtracted.

diagram. While the phase of the theoretical polar motion signal agrees well with the measurements, there remain small deviations in amplitude. Since the model assumptions are based on a simplified deformable Earth one may expect that eventually a better insight into the Earth interior may gained.

However, before this goal can be addressed it is necessary to get a better control of the different bias mechanisms of large ring lasers, which were mentioned several times before. While the diurnal polar motion can be understood as a wobble of the Earth rotation axis, laser gyros also experience tilts from solid Earth tides. These signals occur at a period of half a siderial day and have amplitudes of up to 40 nrad in Wettzell. For the ring lasers in Christchurch one can see additional tilts from ocean loading, which make this effect approximately 3 times larger [21]. Figure 19 shows a spectrum of the ring laser measurements of G taken without interruption over a

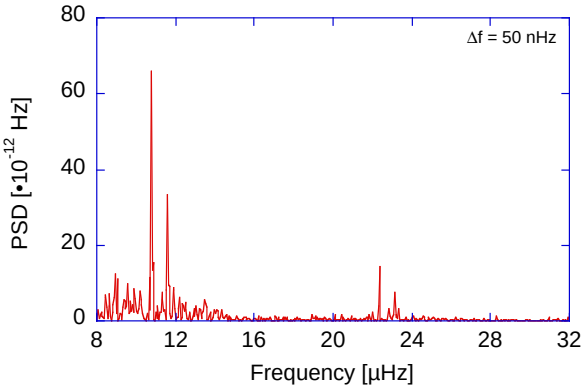


Figure 19. Spectrum of the G ring laser taken from a dataset as long as 243 days. The major signals for diurnal polar motion and solid Earth tides are clearly visible.

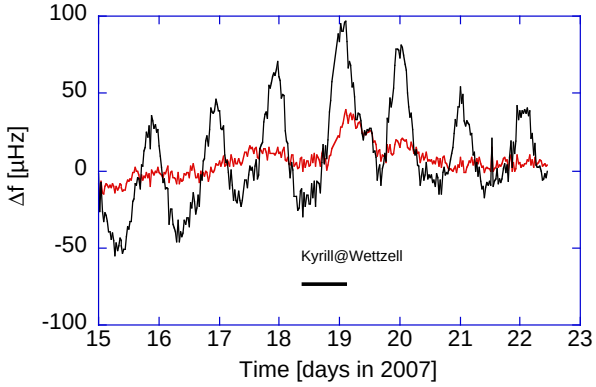


Figure 20. Time series of G measurements taken over a week around the storm "Kyrill" in January 2007. When Earth tides and diurnal polar motion are removed from the data, a distinct transient feature remains in the dataset.

period of 243 days. Both the main contributors to the daily polar motion signals and the solid Earth tides show up clearly. They agree with the literature values to within the spectral resolution of the measurement of 50 nHz.

Apart from these known and expected signals there are also non-periodic signatures in the time series of the ring laser, which cannot be unambiguously identified at this point in time. Figure 20 shows such an example. The displayed dataset was taken over 7 days around the storm "Kyrill", which struck central Europe on the 18th of January, 2007. One can see the diurnal polar motion and solid Earth tides signal, which look very similar to Fig. 18. The second graph in the diagram shows the same dataset with these known components removed. A very distinct signal remains with a maximum on the day after the low pressure area had passed over Europe. Neither the signal itself nor the time delay has been understood so far. For a better illustration the actual time at which the storm passed over the gyroscope is also marked on this plot. Apart from crustal deformation and strain effects due to wind friction causing either tilt or some sort of local rotation also sensor internal artifacts may be responsible for such sensor responses and the investigation of these transient effects is still ongoing.

7 Application in seismology

With the availabilty of large ring lasers geophysical rotations became accessible at a global scale and at various timescales [23]. In particular rotation signals from teleseismic events became measurable for the first time [27, 29]. A specific project on rotational seismology, funded by the German Ministry of Education and Research (BMBF) within the geotechnology program, made the construction of a large ring laser for seismological studies possible. Results from this project eventually led to the

formation of a new working group on rotational seismology.[2] A more detailed report on the application of ring lasers for rotational seismology is given in [26, 34–36].

7.1 Detection of seismic signals

The range of angular velocities to be covered in this application is very wide, i. e. $10^{-14}\,\mathrm{rad/s} \leq \Omega_s \leq 1\,\mathrm{rad/s}$, with the required frequency bandwidth for the seismic waves in the range of $3\,\mathrm{mHz} \leq f_s \leq 10$ Hz [33]. Currently large ring lasers are the only available rotation sensors which meet these demands. Three such devices mounted in orthogonal orientations may eventually provide the quantitative detection of rotations from shear, Love and Rayleigh waves. It is important to note that ring laser gyroscopes are sensitive only to rotations around their area normal vector. At the same time they are completely insensitive to translational motions. From that point of view they provide additional information and may also be very useful to separate between tilt and translations, a persisting problem in seismology.

The goal of the GEOsensor project was the construction and evaluation of a field-deployable demonstrator unit, which eventually will provide access to all six degrees of freedom of motion. The recording of the (complete) earthquake-induced rotational motion is expected to be particularly useful for (1) further constraining earthquake source processes when observed close to the active faults [30]; (2) estimating permanent displacement from seismic recordings [32]; (3) estimating local (horizontal) phase velocities from collocated observations of translations and rotations [26]. Because of the relatively short duration of an earthquake such ring lasers do not need a long-term stability over weeks or months, which is difficult and expensive to obtain. An instrumental stability of approximately one hour during a seismic event is sufficient. Therefore it is possible to use a steel structure attached to a solid concrete platform as the main components of the Sagnac interferometer.

As indicated above, ring lasers for seismic studies require a high data rate of at least 20 Hz, because of the wide bandwidth of seismic frequencies near an earthquake source. While large ring lasers for geodetic applications are usually optimized for measuring variations in the rotation rate of the Earth in a frequency band well below 1 mHz, autoregressive algorithms can be used to determine the Sagnac frequency with a resolution below the Nyquist limit. While this method can still be employed for the strongly bandwidth limited teleseismic signals [28], an entirely different detection scheme is needed for the data evaluation of regional or local seismic events.

Unlike seismometers, the concept of a Sagnac interferometer is not based on mass inertia. As a consequence ring lasers have no moving mechanical parts. This has the advantage that there is no restitution process required for the extraction of the true ground motion from the transfer function of the measurement device. In order to distinguish true measured ground rotations from possible unknown sensor artifacts, two independent ring lasers, namely C-II and UG1 (see Table 1), were collocated and operated at the same place. C-II is placed inside UG1 and the area normal vectors of both ring laser planes are collinear. According to the ring laser equation the relationship between the obtained Sagnac frequency and the input rotation rate

[2] See `http://www.rotational-seismology.org`

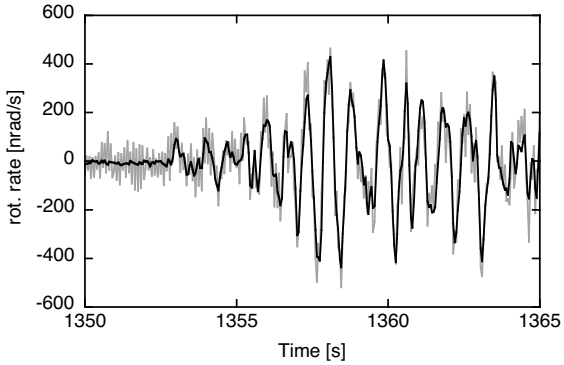


Figure 21. Comparison of two ring laser seismogramms from the the same mag. 7.7 earthquake near Fiji on August 19, 2002. Both ring lasers were located in the same place with identical orientation. Apart from the higher noise level of the smaller instrument the recordings are identical.

is linear over a wide dynamic range. A mag. 7.7 earthquake near the Fiji Islands on August 19, 2002 was recorded simultaneously on both ring lasers. Figure 21 shows the record of the first 15 seconds of this earthquake. The measured raw Sagnac frequency as a function of time was converted to rotation rate in nano-radians per second using Eq. (2). Apart from this conversion the data has not been modified in any way. The dataset from the C-II ring laser is a little noisier than the data from UG1 because there is almost a factor of 20 difference in the respective scale factors. Nevertheless one can see that both ring lasers measure exactly the same signal in phase as well as in amplitude.

Ring lasers provide optical interferograms where the external rate of rotation is proportional to the rate of change of the fringe pattern. This signal becomes available as an audio-frequency at the output of a photomultiplier tube, which is a major difference to the amplitude variations typically recorded by seismometers.

In seismology it is important to detect the rate of change of the Sagnac frequency at 50 ms intervals (20 Hz) very accurately. Since frequency counting techniques do not provide a sufficient resolution at such short integration times, a frequency demodulation concept has been developed. A voltage controlled oscillator is phase locked to the Sagnac frequency of the ring laser, exploiting the fact that Earth rotation provides a constant rate bias in the absence of any seismically induced rotation signals. In the event of an earthquake one obtains the rate of change of the Sagnac frequency at the feedback line of the voltage controlled oscillator. This voltage can be digitized and averaged at the required 20 Hz rate or higher.

Currently the upper limit for the detectable rate of change from a large ring laser is not set by the sensor itself but by the frequency extraction process. To outline the importance of the frequency demodulation technique two earthquakes with distinctly different properties are compared. Figure 22 shows an example for a teleseismic event and an example from a much closer regional earthquake. While for the remote earthquake the spectral power density essentially drops off to zero above frequencies

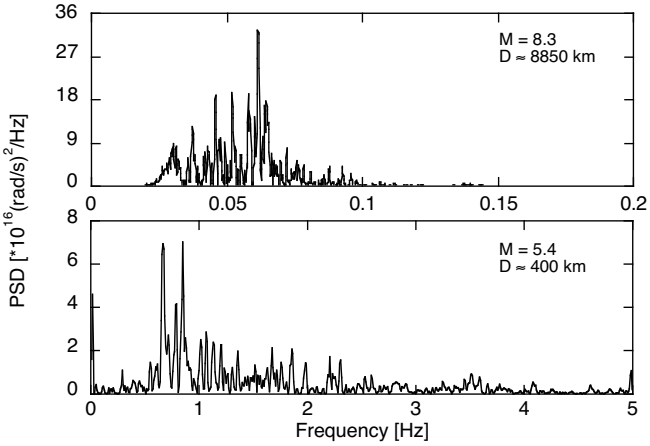


Figure 22. Comparison of recorded rotation spectra from an teleseismic event (Hokkaido: Sept. 9, 2003) and a regional earthquake (France: Feb. 22, 2003). The much higher bandwidth of the rotational wave spectra requires alternative data acquisition techniques such as the demodulator.

of 0.1 Hz, one can still see some signal signature up to about 4 Hz for the regional event. Frequencies with a rate of change above 2 Hz, however, are already outside the regime of reliable representation in phase and amplitude by conventional frequency counting and second-order autoregression frequency analysis [28].

Figure 23 illustrates some basic characteristics for the detection of rotations from seismic signals. The diagram shows most of the measurement range of interest for seismic studies. The relevant frequency window is plotted horizontally, while the

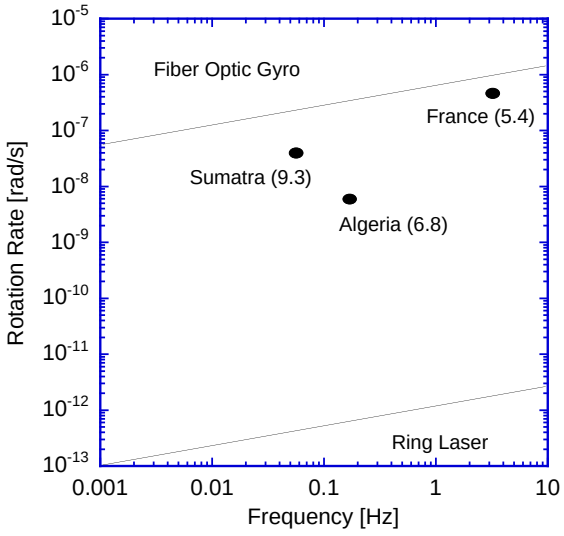


Figure 23. Sensor resolution of different rotation sensor concepts in relation to the observed signal strength of some earthquakes at different epicentral distances.

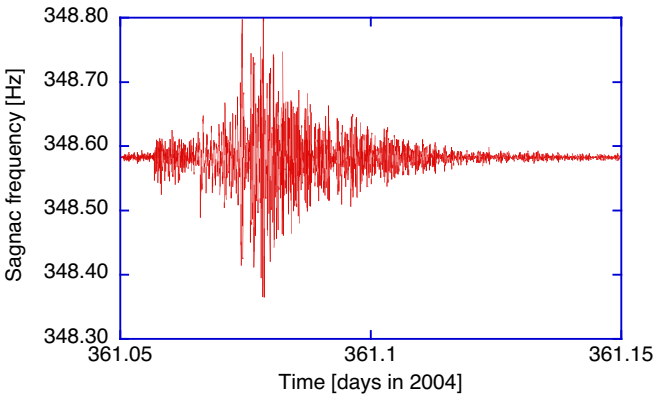


Figure 24. The raw rotation measurement of the mag. 9.3 Sumatra earthquake from December 26, 2004. The dataset was recorded with a good signal to noise ratio.

magnitude of the respective rotation rates is displayed on the vertical. In order to keep this diagram simple the strong motion region is not shown. In the lower part of the plot one can see a line, which indicates the resolution limit for current ring lasers. Depending on the actual scale factor the sensitivity differs from one ring laser to another. However within the scale of this chart this line gives a good representation for the existing large ring lasers in general. Current high quality fibre-optic gyros (FOG) exhibit a sensor resolution of $\delta\varphi = 0.1°/\sqrt{h}$ or slightly less. The upper line was derived from test measurements of a sample FOG type: μFORS-1 manufactured by LITEF GmbH in Germany. Both lines are sloping over the frequency range of interest. This reflects the improvement resulting from longer integration time as the frequency of interest reduces.

To give an idea of the actual sensor requirement, three very different example earthquakes are indicated on the graph. The details of these earthquakes are given in Table 2. Since the Earth crust acts as a lowpass filter one can see that the earthquakes at the right side of the Fig. 23 plot are the closest. All events listed here produced datasets with good signal to noise ratio on the G ring laser. Figure 24 shows the rotational seismogram of the mag. 9.3 Sumatra event as an example. None of these events would have been within the sensor resolution of a FOG at larger distances. As Fig. 23 clearly indicates the application of FOGs for seismic studies is currently only possible for strong motion applications.

source	magnitude	distance
Sumatra	9.3	$>$ 10000 km
Algeria	6.8	1550 km
France	5.4	400 km

Table 2. Details of some earthquakes recorded in Germany by rotational sensors.

7.2 The ring laser component

In order to obtain a stable interferogram of the two laser beams the cavity length has to be kept constant to within a fraction of a wavelength. Therefore usually ring laser bodies are made from Zerodur, a glass ceramic which exhibits a very small relative thermal expansion of $\alpha = 5 \cdot 10^{-8}\,\mathrm{K}^{-1}$. Since a ring laser for seismic applications requires an enclosed area of more than $1\,\mathrm{m}^2$ a monolithic ring construction would both be too expensive and not transportable.

Figure 25 gives an impression of the actually realized ring laser hardware. Again the laser cavity has the shape of a square. The 4 turning mirrors are each located in a solid corner box. As shown on the right side of the plot, a folded lever system allows the alignment of each mirror to within ± 10 seconds of arc. This high level of alignment is required to ensure lasing from an optically stable cavity. The mirrors are located inside steel containers which in turn are connected up with stainless steel tubes, forming a vacuum recipient around the laser beam path. In the middle of one side the steel tubes are reduced to a small glass capillary of 4 mm in diameter and a length of 10 cm, which is required for gain medium excitation. When operated, the ring laser cavity is first evacuated and then filled with a mixture of Helium and Neon reaching a total gas pressure of approximately 6 mbar. The left part of Fig. 25 gives an impression of the instrumental layout. The following two important considerations are unique for the GEOsensor design.

- Since the ring laser is constructed from several components, it requires a stable concrete platform as a base at the location of deployment. Such a pad is simple to specifiy and can be prepared totally independent of the actual GEOsensor deployment.

- The actual area of the ring laser component is not predetermined by the design. The instrument can be built according to the available space at the respective host observatory. Different GEOsensor realizations may therefore have different sizes and consequently different instrumental resolution. The length of the current instrument is 1.6 m on a side, which provides an area of $2.56\,\mathrm{m}^2$.

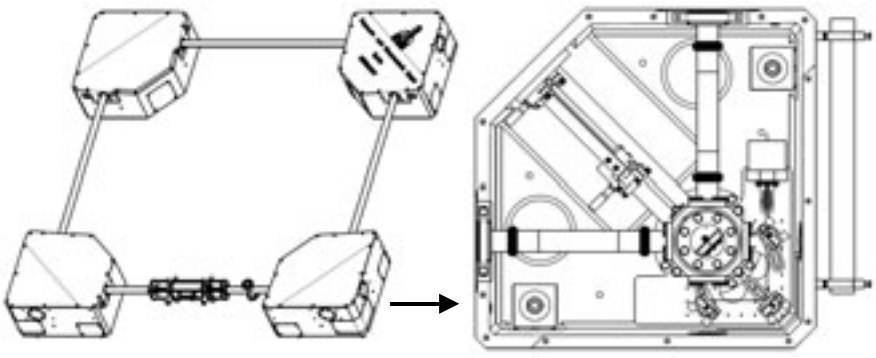


Figure 25. The basic construction layout of the GEOsensor ring laser

location	frequency [Hz]
Wettzell (49.145 N)	138
Pinon Flat (33.6 N)	102
Tokyo (35.4 N)	106
Cashmere (43.57 S)	127

Table 3. Earth rotation bias for some possible GEOsensor locations.

In order to operate the GEOsensor the cavity must be evacuated, baked and filled with a Helium-Neon gas mixture. This procedure requires a turbo molecular pump system and a manifold with a supply of ^{4}He, ^{20}Ne and ^{22}Ne. The pump system is not required during the operation of the GEOsensor but is necessary for the preparation of the instrument and once or twice during a year in order to change the laser gas. Laser excitation itself is achieved via a high-frequency generator, matched to a symmetrical high-impedance antenna at the gain tube. A feedback loop maintains the level of intensity inside the ring laser and ensures monomode operation. When the ring laser is operated it detects the beat note caused by Earth rotation. The magnitude of this beat frequency is depending on $\sin(\Phi)$ with Φ being the latitude of the ring laser location. Table 3 shows the value of the Earth's rate bias for a few locations of interest. Until today the GEOsensor was operated at the first two locations only.

7.3 Deployment of the GEOsensor

After the development of the GEOsensor and including a test installation at the Geodetic Observatory Wettzell, the instrument was shipped to the Scripps Institution of Oceanography in San Diego, California. In January 2005 the installation of the complete sensor took place at the Seismological Observatory Pinon Flat as shown in Fig. 26. The site is located between the San Jacinto and the San Andreas fault. The goal of this installation is the measurement of a number of earthquakes at short distances. This will allow an extensive and systematic study of rotational motions

Figure 26. The ring laser vault under construction at the Pinon Flat (Ca) observatory and the GEOsensor installation in one of the chambers.

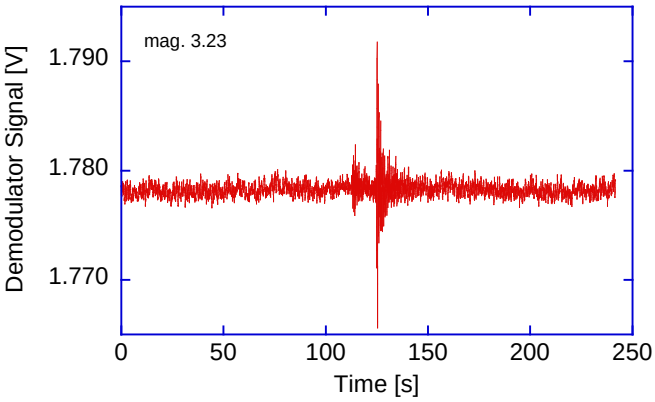


Figure 27. Example of a regional seismic event recorded on the GEOsensor in Pinon Flats.

on all scales with a particular emphasis on local and regional scales with source distances of up to 100 km. From the calculation of theoretical seismograms for all six degrees of freedom of motion, it became apparent that rotational motion information may contribute the most for the investigation of local and regional earthquakes. One example out of many of the recorded seismic events in Pinon Flat is shown in Fig. 27. The basic sensor concept is well suited for the desired application.

7.4 Observations of rotations

Currently there are primarily two types of measurements that are routinely used to monitor global and regional seismic wave fields. Standard inertial seismometers measure three components of translational ground displacement and provide the basis for monitoring seismic activity and ground motion. The second type aims at measuring the deformation of the Earth (strains). It is well known that there is a third type of measurement that should be observed in seismology and geodesy in order to fully describe the motion at a given point, the measurement of ground rotation.

In fact, Aki and Richards [24, 25] have demanded to use rotations for more than 20 years, but due to the lack of suitable sensors this has not been done in the past. The recording of the (complete) earthquake-induced rotational motion is expected to be useful particularly for (1) further constraining earthquake source processes when observed close to the active faults [30]; (2) estimating permanent displacement from seismic recordings [32]; (3) estimating local (horizontal) phase velocities from collocated observations of translations and rotations. As will be shown below, the consistency of broadband ring laser observations of the vertical component of rotation rate observed for distant large earthquakes is readily obtained. Furthermore one can model the observations with numerical simulations of the complete rotational wave field in a three-dimensional heterogeneous global Earth model.

In order to compare translations (measured by a standard seismometer) with the vertical component of the vector of rotation – which is what the G-ring is measur-

ing – the horizontal components of seismic recordings were rotated into radial and transverse directions. Note that Rayleigh waves should not generate such a vertical rotation component, while Love waves are horizontally polarized, hence generate rotations around a vertical axis only. To obtain transverse acceleration, the transverse velocity seismograms were differentiated with respect to time. Under the assumption of a transversely polarized plane wave with displacement $\boldsymbol{u} = (0, u_y(t - x/c), 0)$ and the horizontal phase velocity c, the vector of rotation (curl) is obtained as $\frac{1}{2}\nabla \times \boldsymbol{u} = \left(0, 0, -\frac{1}{2c}\dot{u}_y(t - \frac{x}{c})\right)$ with the corresponding z-component of rotation rate $\Omega_z(x,t) = -\frac{1}{2c}\ddot{u}_y(t - \frac{x}{c})$. This means that at any time rotation rate and transverse acceleration are in phase and the amplitudes are related by $\ddot{u}_y(x,t)/\Omega_z(x,t) = -2c$.

In practise, the phase velocities can be estimated by dividing best-fitting waveforms in sliding a time-window of appropriate length along the seismic signal and rotation rate. Thus, under the plane-wave assumption both signals should be equal in phase and amplitude [27, 29]. This assumption is expected to hold for a considerable part of the observed ground motion due to the large epicentral distance compared to the considered wavelengths and source dimensions. This property is exploited here to verify the consistency of the observations. Close to the seismic source this assumption no longer holds and may form the basis for further constraining rupture processes [30, 31]. A data example (rotation rate and transverse acceleration) of the mag. 8.1 Tokachi-oki event on September 25, 2003, and a time-dependent normalized cross-correlation coefficient (maximum in a 30 s sliding window) is given in Fig. 28. The time window also contains an event (increase in cross-correlation at 3500 s) that was barely visible in the seismograms without correlating the two signals.

When the waveform fit between rotation rate and transverse acceleration is sufficiently good (e. g., a normalized correlation coefficient >0.95), one can estimate phase velocities by dividing the peak amplitudes of both traces as explained above. These time dependent estimates of phase velocities are shown in Fig. 28 (bottom). Despite the scattering of the phase velocities in the time window containing the Love waves (9000-10000 s and around 14000 s for the aftershock) the estimates are in the right range of expected phase velocities and the negative slope of the velocities with time suggest that the expected dispersive behavior (earlier longer periods have higher phase velocities) can be observed in the data using this processing approach. The lack of correlation in the time windows excluding the Love waves may indicate that either the (body-) wave fronts are not planar or that the energy comes from out-of-plane directions through scattering. It should be noted, that such results are obtained constistently.

8 Summary

Large ring lasers have come a long way over the last decade. In the 1994 the construction of C-II was met with a lot of skepticism. From the initial sensor stability of only 0.1% with respect to Earth rotation, C-II moved quickly to a domain of better than 1 part in 10^6. In 1997 a concept for an even larger ring laser, G, was proposed for the first time. Then it was not even known if such a cavity would allow single mode operation at all, a prerequisite for the operation of a Sagnac interferometer.

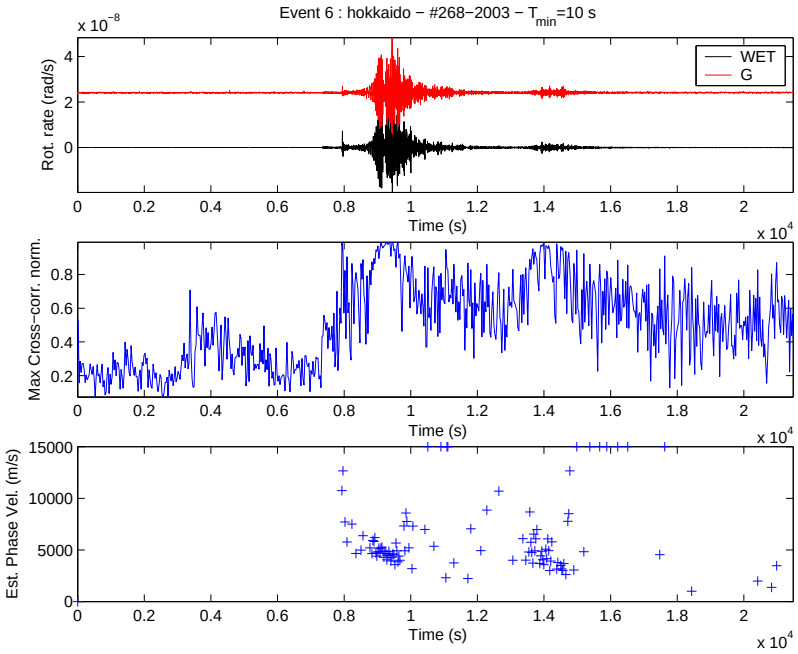


Figure 28. Top: Observation of rotation rate (upper trace) and transverse acceleration (lower trace) after the mag. 8.1 Tokachi-oki event, 29-9-03. Middle: The cross-correlation-coefficient in a 30 s sliding window. Note the increase in correlation during the main shock (>7500 s) and aftershock (>13800 s) to almost 1 (perfect match). Bottom: Estimates of horizontal phase velocities in time windows with good phase match (> 9500 s). Note the decreasing phase velocities in the Love wave train (e. g., 8000-10000 s, indicative of Love wave dispersion). This diagram was kindly provided by H. Igel.

This led to the construction of the first cheap stainless steel construction, namely G0. The only objective was to achieve mono-mode operation. However to the surprise of everybody, G0 could be operated as a gyroscope comparatively easily. This achievement laid the foundations for G, the most sophisticated and well designed ring laser of all, shifting the limit of gyroscopic sensor resolution well beyond 1 part in 10^8. Encouraged by this result larger and larger ring lasers were built, demonstrating their viability for the application in this new field of high resolution Sagnac spectroscopy. A lot of technical problems had to be overcome, which in hindsight appeared to be obvious. Today it seems that G has reached the limit of sensitivity which is practically possible for a gyroscope attached to the Earth. The sensor stability has been improved from stable operations at the scale of minutes about 10 years ago up to many days today. As this lane is pursued further more and more geophysical signals with longer periods are expected to become visible, which eventually may make significant contributions to the field of space geodesy. Furthermore a number of exciting results from the studies of rotational seismology are expected from Sagnac gyroscopes.

Acknowledgements. This project review could have had more than six co-authors easily. However that was impractical to coordinate in the available time. I have to take responsibility for the statements made and and the balancing of the items presented. Appropriate reference is given to earlier publications covering the various aspects of the ring laser project in more detail. The combined ring laser results were possible because of a collaboration of Forschungseinrichtung Satellitengeodäsie, Technische Universität München, Germany, University of Canterbury, Christchurch, New Zealand Bundesamt für Kartographie und Geodäsie, Frankfurt, Germany and the Ludwig-Maximilian-Universität, München, Germany. The GEOsensor was funded under the program GEOTECHNOLOGIEN of BMBF and DFG, Grant 03F0325 A-D. University of Canterbury research grants, contracts of the Marsden Fund of the Royal Society of New Zealand and also grants from the Deutsche Forschungsgemeinschaft (DFG) are gratefully acknowledged. Special thanks go to my colleagues Heiner Igel, Geoffrey Stedman, Robert Hurst, Jon-Paul Wells, Clive Rowe, Robert Thirkettle, Thomas Klügel, Alexander Velikoseltsev, Manfred Schneider, Benedikt Pritsch, Wolfgang Schlüter and Robert Dunn for their contributions.

References

[1] G. Sagnac, Comptes Rendus, Acac. Sci. (Paris) **157**, 708 (1913).
[2] B. Höling, *Ein Lasergyroskop zur Messung der Erdrotation*, Dissertation, Universität Tübingen (1990).
[3] P. Milonni and J. Eberly, *Lasers* (Wiley & Sons, New York, 1988), pp. 589.
[4] G. E. Stedman, 'Ring-laser tests of fundamental physics and geophysics', Rep. Prog. Phys. **60**, 615 (1997).
[5] A. A. Michelson and H. G. Gale, 'The Effect of the Earth's Rotation on the Velocity of Light', Astrophys. J. **61**, 140 – 145 (1925).
[6] D. Loukianov, R. Rodloff, H. Sorg, and B. Stieler (Eds.), *RTO AGARDograph 339, Optical Gyros and their Application* (NATO RTO, 1999), ISBN 92-837-1014-2.
[7] K. U. Schreiber, *Ringlasertechnologie für geowissenschaftliche Anwendungen*, Habilitationsschrift, Mitteilungen d. Bundesamtes f. Kartographie und Geodäsie, Bd. 8 (1999).
[8] H. Bilger, G. E. Stedman, Z. Li, K. U. Schreiber, and M. Schneider, 'Ring Lasers for Geodesy', IEEE Trans. Instr. Meas. **44**, 468 (1995).
[9] F. Aronowitz, 'The laser gyro', in *Laser applications, Vol. 1*, edited by M. Ross (Academic Press, New York, 1971), pp. 133 – 200.
[10] F. Aronowitz, 'Fundamentals of the Ring Laser Gyro', in *RTO AGARDograph 339, Optical Gyros and their Application*, edited by D. Loukianov, R. Rodloff, H. Sorg, and B. Stieler (NATO RTO, 1999).
[11] J. R. Wilkinson, 'Ring Lasers', Prog. Quant. Electr. **11**, 1 (1987).
[12] A. Velikoseltsev, *The development of a sensor model for Large Ring Lasers and their application in seismic studies*, Dissertation, Technische Universität München (2005).
[13] U. E. Hochuli, P. Haldemann, H. A. Li, 'Factors influencing the relative frequency stability of He-Ne laser structures', Rev. Sci. Instr. **45**, 1378 (1974).
[14] W. E. Ahearn and R. E. Horstmann, 'Nondestructive Analysis for HeNe Lasers', IBM J. Res. Develop. **23**, 128 (1979).
[15] W. W. Chow, J. Gea-Banacloche, L. M. Pedrotti, V. E. Sanders, W. Schleich, and M. O. Scully, 'The ring laser gyro', Rev. Mod. Phys. **57**, 61 (1985).
[16] J. T. Verdeyen, *Laser Electronics*, 3rd ed. (Prentice Hall, London, 2000).
[17] P. McClure, *Diurnal polar motion*, GSFC Rep. X-529-73-259, (Goddard Space Flight Center, Greenbelt, Md., 1973).

[18] V. Frede and V. Dehant, 'Analytical verses semi-analytical determinations of the Oppolzer terms for a non-rigid Earth', J. Geodesy **73**, 94 (1999).
[19] H. Moritz and I. I. Mueller, *Earth Rotation. Theory and Observation* (Ungar Publishing Co., New York, 1987).
[20] A. Brzeziński, 'Contribution to the theory of polar motion for an elastic earth with liquid core', Manuscripta Geodaetica **11**, 226 (1986).
[21] K. U. Schreiber, G. E. Stedman, and T. Klügel, 'Earth tide and tilt detection by a ring laser gyroscope', J. Geophys. Res. **108(B2)**, 2132 (2003), doi:10.1029/2001JB000569.
[22] K. U. Schreiber, A. Velikoseltsev, M. Rothacher, T. Klügel, G. E. Stedman, and D. L. Wiltshire, 'Direct measurement of diurnal polar motion by ring laser gyroscopes', J. Geophys. Res. **109**, B06405 (2004), doi:10.1029/2003JB002803.
[23] K. U. Schreiber, M. Schneider, C. H. Rowe, G. E. Stedman, and W. Schlüter, 'Aspects of Ring Lasers as Local Earth Rotation Sensors', Surveys in Geophysics **22**, 603 (2001).
[24] K. Aki and P. G. Richards, *Quantitative seismology*, 1st ed. (Freeman and Company, New York, 1980)
[25] K. Aki, P. G. Richards, *Quantitative seismology*, 2nd ed. (University Science Books, Sausalito, CA, 2002).
[26] H. Igel, K. U. Schreiber, B. Schuberth, A. Flaws, A. Velikoseltsev, and A. Cochard, 'Observation and modelling of rotational motions induced by distant large earthquakes: the M8.1 Tokachi-oki earthquake September 25, 2003', Geophys. Res. Lett. **32**, L08309 (2005), doi:10.1029/2004GL022336.
[27] D. P. McLeod, G. E. Stedman, T. H. Webb, and K. U. Schreiber, 'Comparison of standard and ring laser rotational seismograms', Bull. Seism. Soc. Amer. **88**, 1495 (1998).
[28] D. P. McLeod, B. T. King, G. E. Stedman, K. U. Schreiber, and T. H. Webb, 'Autoregressive analysis for the detection of earthquakes with a ring laser gyroscope', Fluctuations and Noise Letters **1**, R41 (2001).
[29] A. Pancha, T. H. Webb, G. E. Stedman, D. P. McLeod, and K. U. Schreiber, 'Ring laser detection of rotations from teleseismic waves', Geophys. Res. Lett. **27**, 3553 (2000).
[30] M. Takeo and H. M. Ito, 'What can be learned from rotational motions excited by earthquakes?', Geophys. J. Int. **129**, 319 (1997).
[31] M. Takeo, 'Ground rotational motions recorded in near-source region of earthquakes', Geophys. Res. Lett. **25**, 789 (1998).
[32] M. D. Trifunac and M. I. Todorovska, 'A note on the usable dynamic range of accelerographs recording translation', Soil Dyn. Earth. Eng. **21**, 275 (2001).
[33] K. U. Schreiber, A. Velikoseltsev, G. E. Stedman, R. B. Hurst, and T. Klügel, 'Large Ring Laser Gyros as High Resolution Sensors for Applications in Geoscience', in *Proceedings of the 11th International Conference on Integrated Navigation Systems, St. Petersburg* (2004), pp. 326 – 331.
[34] K. U. Schreiber, H. Igel, A. Velikoseltsev, A. Flaws, B. Schuberth, W. Drewitz, and F. Müller, 'The GEOsensor Project: Rotations - a New Observable for Seismology', in *Observation of the Earth System from Space* (Springer, 2005), pp. 427 – 447.
[35] K. U. Schreiber, G. E. Stedman, H. Igel, and A. Flaws, 'Rotational Motions in Seismology: Theory, Observation, Simulation', in *Earthquake Source Asymmetry, Structural Media and Rotational Effects*, edited by R. Teisseyre, M. Takeo, and E. Majewski (Springer, New York, 2006), Chap. 29.
[36] A. Cochard, H. Igel, B. Schuberth, W. Suryanto, A. Velikoseltsev, K. U. Schreiber, J. Wassermann, F. Scherbaum, and D. Vollmer, 'Rotational Motions in Seismology: Theory, Observation, Simulation', in *Earthquake Source Asymmetry, Structural Media and Rotational Effects*, edited by R. Teisseyre, M. Takeo, and E. Majewski (Springer, New York, 2006), Chap. 30.

Oscillations, Waves and Interactions, pp. 311–332
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-12-0

Charge-ordering phenomena in one-dimensional solids

Martin Dressel

1. Physikalisches Institut, Universität Stuttgart
Pfaffenwaldring 57, 70550 Stuttgart, Germany

Email:dressel@pi1.physik.uni-stuttgart.de

Abstract. As the dimensionality is reduced, the world becomes more and more interesting; novel and fascinating phenomena show up which call for understanding. Physics in one dimension is a fascinating topic for theory and experiment: for the former often a simplification, for the latter always a challenge. Various ways will be demonstrated how one-dimensional structures can be achieved in reality. In particular organic conductors could establish themselves as model systems for the investigation of the physics in reduced dimensions; they also have been subject of intensive research at the Dritte Physikalische Institut of Göttingen University over several decades.
In the metallic state of a one-dimensional solid, Fermi-liquid theory breaks down and spin and charge degrees of freedom become separated. But the metallic phase is not stable in one dimension: as the temperature is reduced, the electronic charge and spin tend to arrange themselves in an ordered fashion due to strong correlations. The competition of the different interactions is responsible for which broken-symmetry ground state is eventually realized in a specific compound and which drives the system towards an insulating state.
Here we review the various ordering phenomena and how they can be identified by dielectric and optic measurements. While the final results might look very similar in the case of a charge density wave and a charge-ordered metal, for instance, the physical cause is completely different. When density waves form, a gap opens in the electronic density-of-states at the Fermi energy due to nesting of the one-dimension Fermi-surface sheets. When a one-dimensional metal becomes a charge-ordered Mott insulator, on the other hand, the short-range Coulomb repulsion localizes the charge on the lattice sites and even causes certain charge patterns.

1 Introduction

Although the world is three-dimensional in space, physics in one dimension has always attracted a lot of attention. One-dimensional models are simpler compared to three-dimensional ones, and in many cases can be solved analytically only then [1]. Often the reduction of dimension does not really matter because the essential physics remains unaffected. But there are also numerous phenomena in condensed matter which only or mainly occur in one dimension. In general, the dominance of the lat-

tice is reduced and electronic interactions become superior. Quantum mechanical effects are essential as soon as the confinement approaches the electronic wavelength. Fundamental concepts of physics, like the Fermi liquid theory of interacting particles break down in one dimension and have to be replaced by alternative concepts based on collective excitations [2]. The competition of different interactions concerning the charge, spin, orbital and lattice degrees of freedom can cause ordering phenomena, i. e. phase transitions to a lower-symmetry state as a function of temperature or some order parameter. In one dimension, fluctuations strongly influence the physical properties and smear out phase transitions. An interesting task now is to approximate one-dimensional systems in reality and check the theoretical predictions. Besides pure scientific interest, the crucial importance of these phenomena in nanotechnology might not lie too far ahead.

2 Realization of one-dimensional systems

2.1 Artificial structures

The ideal one-dimensional system would be an infinite chain of atoms in vacuum; close enough to interact with their neighbours, but completely isolated from the environment. Over the past years, significant progress has been made towards the realization of one-dimensional atomic gases, based on Bose-Einstein condensates of alkalides trapped in two-dimensional optical lattices [3]; however, besides other severe drawbacks, only a limited number of investigations can be performed on quantum gases in order to elucidate their properties.

In solids one-dimensional physics can be achieved in various ways. The most obvious approach would be to utilize semiconductor technology. There layers can be prepared by atomic precision, using molecular beam epitaxy that leads to a two-dimensional electron gas at interfaces and quantum wells [4]. Employing electron-beam lithography and advanced etching technology, one-dimensional quantum wires are fabricated with an effective width comparable to the wavelength of the electrons (Fig. 1). Besides the enormous technological effort, the disadvantage of this approach is that these structures are embedded in bulk materials and not easily accessible to further experiments.

If the surface of a single crystal, like silicon, is cut in a small angle with respect to a crystallographic direction, terraces are produced on the surface with mono-

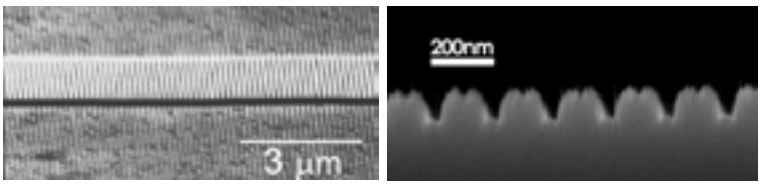


Figure 1. One-dimensional semiconductor quantum wells for GaN lasers (electron micrographs provided by H. Schweizer, Stuttgart). (a) The ridge waveguide covers an area of $1000\,\mu\mathrm{m}\times 6\,\mu\mathrm{m}$; (b) the second-order grating has a period of 190 nm.

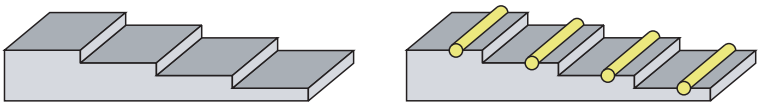


Figure 2. Realization of metallic nanowires: The silicon surface is cut in a certain angle leading to single atomic steps; the width of the terrace depends on the angle. Evaporated gold assembles itself in such a way that atomic wires are formed along the steps.

atomic steps separating them. The surface reconstruction may lead to an anisotropic arrangement with the possibility of one-dimensional structures. When a metal, like gold, is evaporated on top of it, the atoms will organize themselves in rows along these steps as visualized in Fig. 2. Taking care of the surface reconstruction and of the right density of gold eventually leads to chains of gold atoms separated by the terrace width [5]. This is a good model of a one-dimensional metal which can be produced in large quantities, easily covering an area of $1 \times 1\,\mathrm{cm}^2$. As common in surface technology, ultra-high vacuum is required, and only in situ experiments – like electron diffraction, tunnelling or photoemission spectroscopy – have been performed by now.

One-dimensional topological defects in single crystals, known as dislocations, are an intriguing possibility to achieve a one-dimensional metal, which was utilized by H.-W. Helberg and his group [6] in the frame of the Göttinger Sonderforschungsbereich 126. Dislocations in silicon consist of chains of Si atoms, each having a dangling bond as depicted in Fig. 3, i. e. a non-saturated half-filled orbital [7]. Along these rows, metallic conduction is possible while in the perpendicular direction they are isolated. Since dc measurements with microcontacts on both ends of a single dislocation are challenging, contactless microwave experiments were developed as the prime tool to investigate the electronic transport along dislocations in silicon and germanium [6].

It is possible to grow bulk materials as extremely thin and long hair-like wires when stress is applied; they are known as whiskers of gold, silver, zinc, tin, etc.

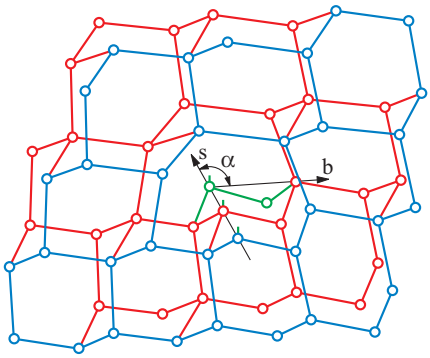


Figure 3. 60° dislocation in a (111) plane of a diamond lattice, the Burgers vector points in the direction of **b**. At the edge of the additional plane (indicated by **s**) the covalent bonds have no partner.

Metallic whiskers often lead to circuit shortages and failures, and are sought to be avoided. An enormous potential of applications is seen in another sort of filaments solely consisting of carbon atoms: carbon nanotubes. They can be considered as rolled-up sheets of graphite, with electrical properties very much depending on the winding ratio. Single-wall carbon nanotubes with a small diameter and the right winding ratio are excellent realizations of one-dimensional conductors [8].

2.2 Anisotropic crystals

By far the most successful approach to one-dimensional physics are highly anisotropic crystals. Here $K_2Pt(CN)_4Br_{0.3}{\cdot}H_2O$, known as KCP, represents the most intuitive example which consists of a chain of platinum ions with overlapping d orbitals, as depicted in Fig. 4(a). The Pt separation is only 2.894 Å along the chain direction while the distance between the chains is 9.89 Å. The Br counterions remove electrons from the planar $Pt(CN)_4$ units and the resulting fractional charge $Pt^{1.7}(CN)_4$ leads to a partially filled electron band, the prerequisite for metallic behaviour. The room temperature conductivity along the chain direction is very high $\sigma_{\parallel} = 10^2\,(\Omega\mathrm{cm})^{-1}$. The anisotropy ratio is $\sigma_{\parallel}/\sigma_{\perp} = 10^5$ [9].

Transition metal oxides are known for decades to form low-dimensional crystal structures [10]. Varying the composition and structural arrangement provides the possibility to obtain one- and two-dimensional conductors or superconductors, but also spin chains and ladders. The interplay of the different degrees of freedom together with the importance of electronic correlations makes these systems an almost

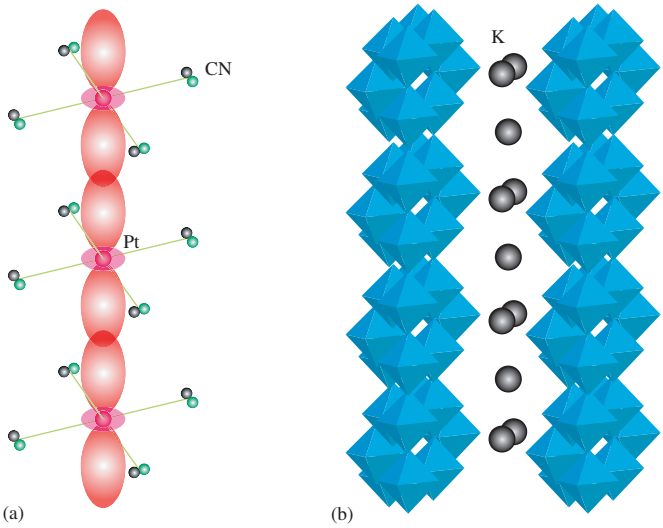


Figure 4. (a) In $K_2Pt(CN)_4Br_{0.3}{\cdot}H_2O$ (KCP) the platinum ions form chains of overlapping orbitals, leading to a metallic conductivity. (b) Sharing edges and corners, the molybdenum oxide octahedra in $K_{0.3}MoO_3$ (blue bronze) form chains along the b direction. Alkali-ions like K or Rb provide the charge.

unlimited source for novel and exciting phenomena and a challenge for their theoretical understanding [11]. The blue bronze $K_{0.3}MoO_3$ and related compounds established themselves quickly as model systems to study electronic properties of quasi-one-dimensional metals above and below the Peierls transition at $T_{CDW} = 185\,K$ (Fig. 4(b)).

While in KCP the metallic properties are due to the platinum ions, organic conductors form a class of solids with no metal atoms present (or relevant); instead the π electrons distributed over of the entire organic molecule form the orbitals which might overlap and lead to band-like conductivity. The additional degree of freedom, tailoring these molecules, supplements the structural arrangement in the crystal and makes it possible to fine-tune competing contributions for the desired properties. This makes organic materials superior for studying low-dimensional physics and ordering phenomena in solids. Low-dimensional organic crystals were explored at the Drittes Physikalisches Institut of Göttingen University already in the 1970s and 1980s; thus in the following we will constrain ourselves mainly to these examples.

In the course of the last two decades, in particular the Bechgaard salts tetramethyl-tetraselenafulvalene (TMTSF), and its variant TMTTF where selenium is replaced by sulfur, turned out to be an excellent model for quasi-one-dimensional metals, superconductors, charge order, spin-density-wave systems, spin chains, spin-Peierls systems, etc. depending on the degree of coupling along and perpendicular to the chains [12]. The planar organic molecules stack along the a-direction with a distance of approximately 3.6 Å. In the b-direction the coupling between the chains is small, and in the third direction the stacks are even separated by the inorganic anion, like PF_6^-, SbF_6^-, ClO_4^-, Br^-, etc. as depicted in Fig. 5. Each organic molecule transfers half an electron to the counterions. In general, a small dimerization leads to pairs of organic molecules. In addition, spontaneous charge disproportionation, called charge ordering (CO), may divide the molecules into two non-equivalent species (cf. Fig. 11) commonly observed in TMTTF salts. Due to the instability of the quasi one-dimensional Fermi surface, at ambient pressure $(TMTSF)_2PF_6$ undergoes a transition to a spin-density-wave (SDW) ground state at $T_{SDW} = 12\,K$ (cf. Fig. 6). Applying pressure or replacing the PF_6^- anions by ClO_4^- leads to a stronger coupling in the second direction: the material becomes more two-dimensional. This seems to be a requirement for superconductivity as first discovered in 1979 by Jérome and coworkers [12,13].

3 Ordering phenomena

One-dimensional structures are intrinsically instable for thermodynamic reasons. Hence various kinds of ordering phenomena can occur which break the translational symmetry of the lattice, charge or spin degrees of freedom. On the other hand, fluctuations suppress long-range order at any finite temperature in one (and two) dimension. Only the fact that real systems consist of one-dimensional chains, which are coupled to some degree, stabilizes the ordered ground state. The challenge now is to extract the one-dimensional physics from experimental investigations of quasi-one-dimensional systems.

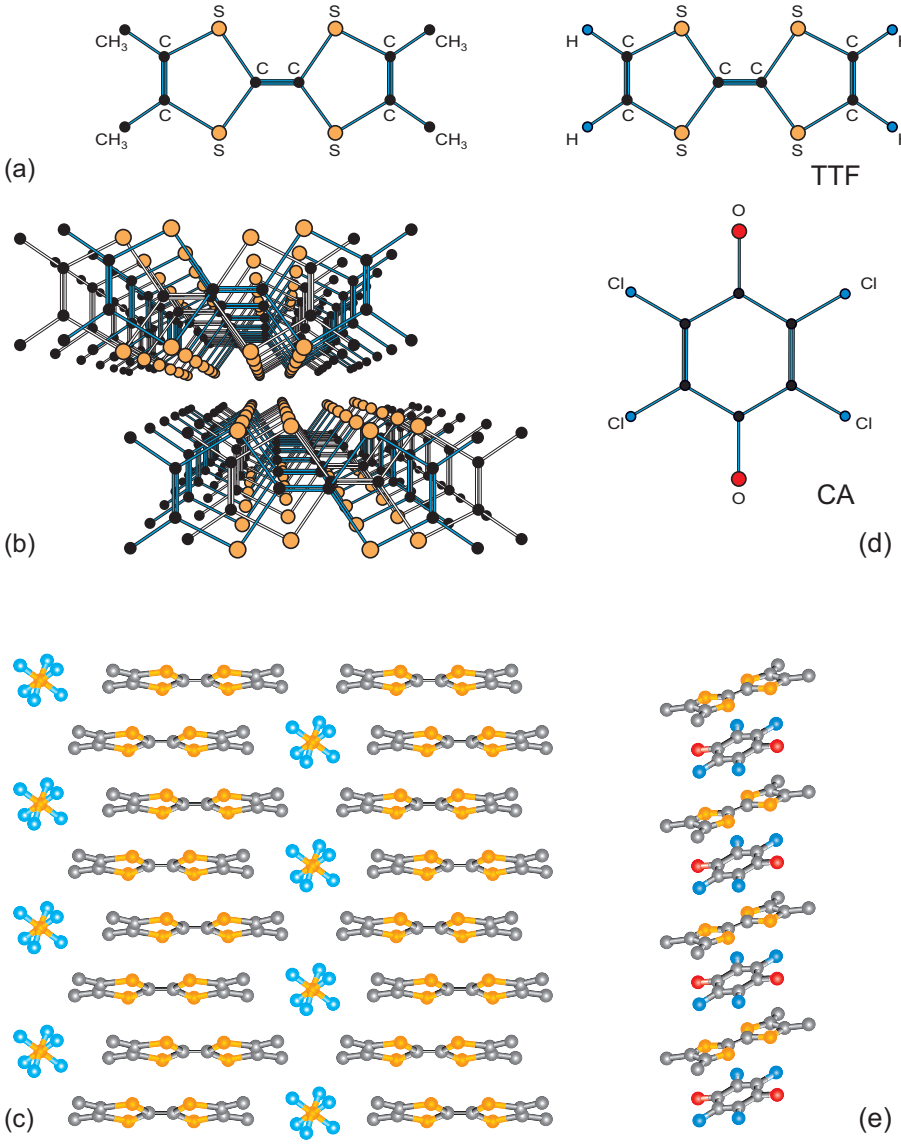


Figure 5. (a) Planar TMTTF molecule. (b) View along the stacks of TMTTF (a-direction) and (c) perpendicular to them (b-direction). Along the c-direction the stacks of the organic molecules are separated by monovalent anions, like PF_6^- or AsF_6^-.
(d) TTF molecule and chloranil QCl_4 (e) in the mixed-stack compound TTF-CA, the planar TTF and CA molecules alternate.

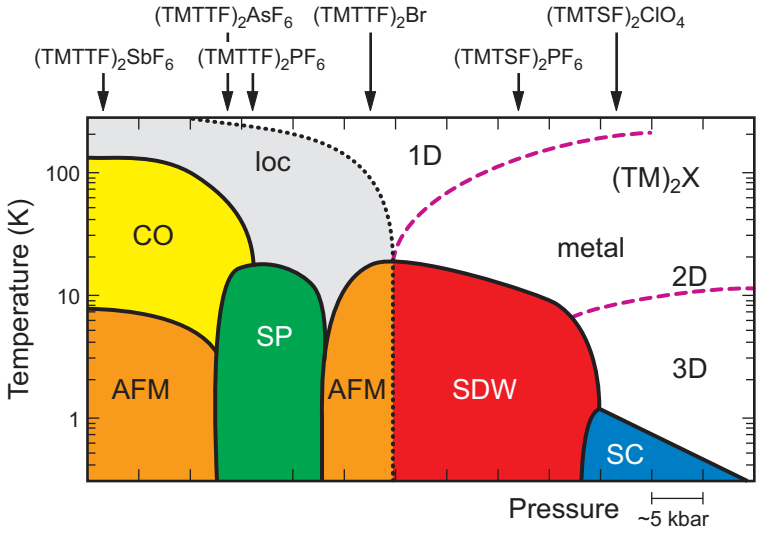


Figure 6. The phase diagram of the quasi one-dimensional TMTTF and TMTSF salts. For the different compounds the ambient-pressure position in the phase diagram is indicated. Going from the left to the right by physical or chemical pressure, the materials get less one-dimensional due to the increasing interaction in the second and third direction. Here loc stands for charge localization, CO for charge ordering, SP for spin-Peierls, AFM for antiferromagnet, SDW for spin density wave, and SC for superconductor. The description of the metallic state changes from a one-dimensional Luttinger liquid to a two- and three-dimensional Fermi liquid. While some of the boundaries are clear phase transitions, the ones indicated by dashed lines are better characterized as a crossover.

At first glance, there seems to be no good reason that in a chain of molecules the sites are not equivalent, or that the itinerant charges of a one-dimensional metal are not homogeneously distributed. However, the translational symmetry can be broken if electron-phonon interaction and electron-electron interaction become strong enough. Energy considerations then cause a redistribution in one or the other way, leading to charge density waves or charge order. Indeed, these ordering phenomena affect most thermodynamic, transport and elastic properties of the crystal; here we want to focus on the electrodynamic response, i.e. optical properties in a broad sense.

First of all, there will be single-particle electron-hole excitations which require energy of typically an eV. But in addition, collective modes are expected. There is a rather general argument by Goldstone [14] that whenever a continuous symmetry is broken, long-wavelength modulations in the symmetry direction should occur at low frequencies. The fact that the lowest energy state has a broken symmetry means that the system is stiff: modulating the order parameter (in amplitude or phase) will cost energy. In crystals, the broken translational order introduces a rigidity to shear deformations, and low-frequency phonons. These collective excitations are expected well below a meV.

3.1 Charge density wave

The energy dispersion forms electronic bands which are filled up to the Fermi wavevector $\mathbf{k}_F$. In one dimension, the Fermi surface consists of only two sheets at $\pm k_F$. The crucial point is that the entire Fermi surface can be mapped onto itself by a $2k_F$ translation. Since the density of states in one dimension diverges as $(E - E_0)^{-1/2}$ at the band-edge E_0, the electronic system is very susceptible to $2k_F$ excitations. The result of the Fermi surface nesting and divergency of the electronic density of states is a spatial modulation in the charge density $\rho(\mathbf{r})$ with a period of $\lambda = \pi/k_F$ (Fig. 7), which does not have to be commensurate to the lattice: this is called a charge density wave (CDW). Long-range charge modulation is crucial because a CDW is a k-space phenomenon. Mediated by electron-phonon coupling, this causes a displacement of the underlying lattice (Peierls instability). The gain in electronic energy due to the lowering of the occupied states has to over-compensate the energy required to modulate the lattice [10,15].

The consequence of the CDW formation is an energy gap 2Δ in the single-particle excitation spectrum, as observed in the activated behaviour of electronic transport or a sharp onset of optical absorption. Additionally, collective excitations are possible which lead to translation of the density wave as a whole. Although pinning to lattice imperfections prevents Fröhlich superconductivity, the density-wave ground state exhibits several spectacular features, like a pronounced non-linearity in the charge transport (sliding CDW) and a strong oscillatory mode in the GHz range of frequency (pinned-mode resonance) [15,16]. In 1974 this behaviour was observed for the first time in the optical properties of KCP [9], but later recovered in all CDW systems. In Fig. 8 the optical reflectivity and conductivity of KCP is displayed for different temperatures and polarizations. Due to the anisotropic nature, the reflectivity $R(\omega)$ shows a plasma edge only for the electric field $\mathbf{E}$ along the chains while it remains low and basically frequency independent perpendicular to it, as known

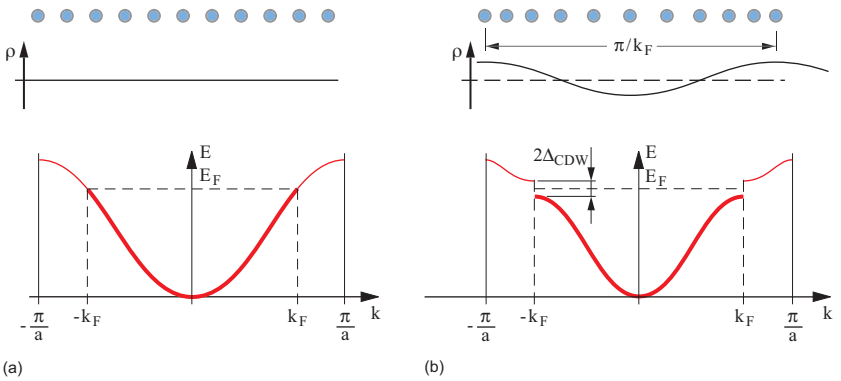


Figure 7. (a) In a regular metal, the charge is homogeneously distributed in space. The conduction band is filled up to the Fermi energy E_F. (b) A modulation of the charge density with a wavelength $\lambda = \pi/k_F$ changes the periodicity; hence in k-space the Brillouin zone is reduced which causes a gap $2\Delta_{\mathrm{CDW}}$ at $\pm k_F$. The system becomes insulating.

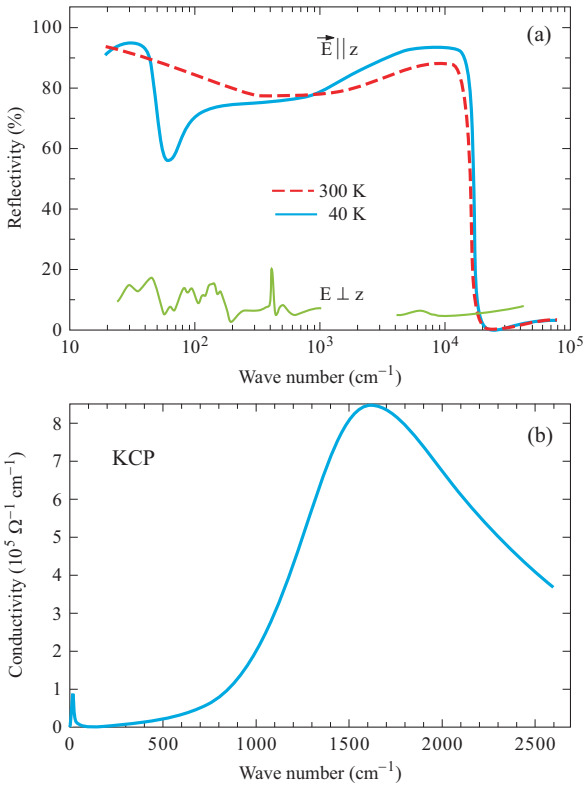


Figure 8. (a) Reflectivity of $K_2Pt(CN)_4Br_{0.3}\cdot H_2O$ (abbreviated KCP) measured parallel and perpendicular to the chains at different temperatures as indicated. (b) Optical conductivity of KCP for $\mathbf{E} \parallel$ stacks at $T = 40\,\mathrm{K}$ (after Ref. [9]). The excitations across the single-particle Peierls gap lead to a broad band in the mid-infrared while the small and sharp peak centered around $15\,\mathrm{cm}^{-1}$ is due to the pinned mode.

from dielectrics. At low temperatures, the single particle gap around $1000\,\mathrm{cm}^{-1}$ becomes more pronounced, and an additional structure is observed in the far-infrared conductivity which is assigned to the pinned-mode resonance induced by the CDW (Fig. 8(b)).

A detailed investigation of the pinned-mode resonance, its center frequency and lineshape, and furthermore its dependence on temperature and impurity content turned out to be extremely difficult because it commonly occurs in the range of 3 to 300 GHz (0.1 to $10\,\mathrm{cm}^{-1}$); i. e. it falls right into the gap between high-frequency experiments using contacts and optical measurements by freely travelling waves [16]. Microwave technique based on resonant cavities and quasioptical THz spectroscopy was advanced over the years in order to bridge this so-called THz gap [17]. Enclosed resonators have been utilized for decades at the Drittes Physikalisches Institut [18] and were readily available when in 1971 I. Shchegolev suggested them as a tool for investigating small and fragile low-dimensional organic crystals like TTF-TCNQ [19].

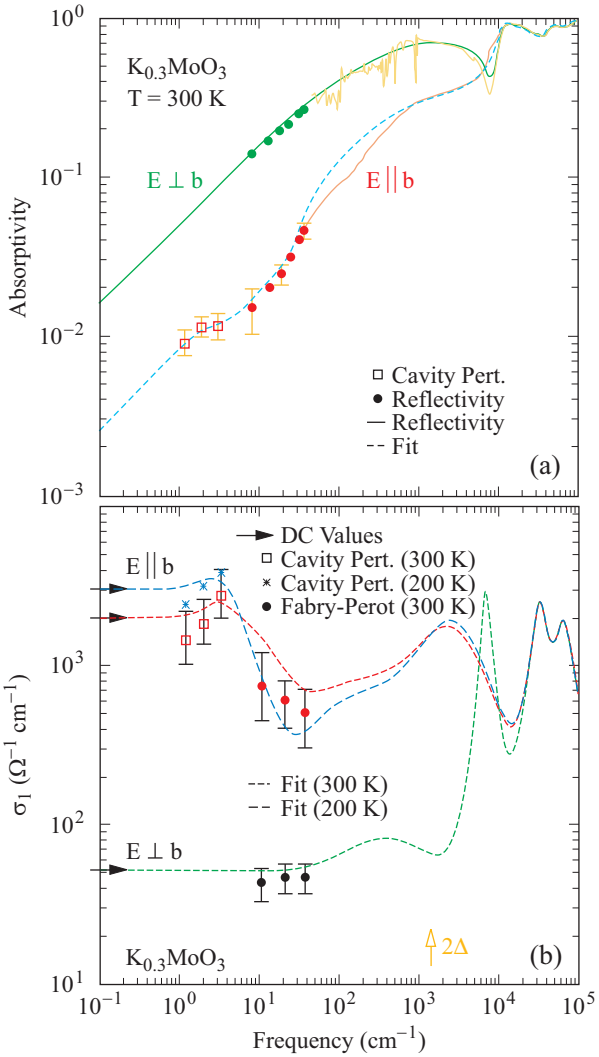


Figure 9. (a) Frequency dependence of the room temperature absorptivity $A = 1 - R$ of blue bronze ($K_{0.3}MoO_3$) in both orientations **E** ∥ stacks and **E** ⊥ stacks. The squares were obtained by measuring the surface resistance using cavity perturbation method, the circles represent data of quasioptical reflectivity measurements employing a Fabry-Perot resonator. The solid lines show the results of the dispersion analysis of the data. (b) Optical conductivity of $K_{0.3}MoO_3$ measured parallel and perpendicular to the stacks by standard dc technique (arrows), cavity perturbation (open squares), coherent-source THz spectroscopy (solid dots) and infrared reflectivity. The open arrow indicates the single-particle gap as estimated from dc measurements below T_{CDW} (after Ref. [20]).

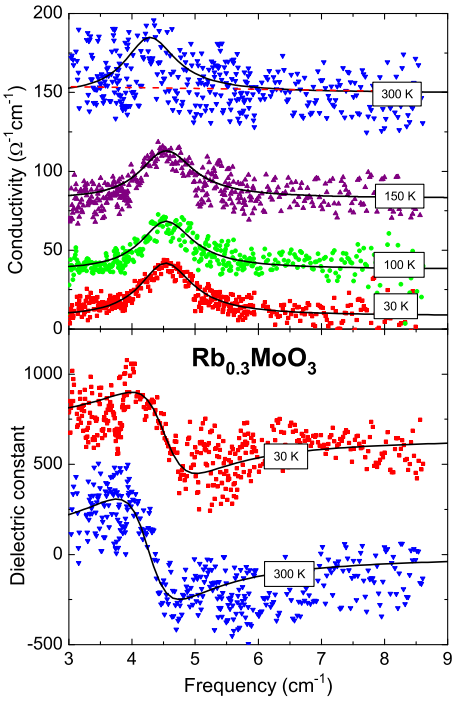


Figure 10. Optical conductivity and dielectric constant of $Rb_{0.3}Mo_3$ at various temperatures above and below T_{CDW} as indicated; note the curves are *not* shifted. The points represent results directly calculated from the transmission and phase-shift spectra. The solid lines correspond to fits (after Ref. [21]). Around $\omega_0/2\pi c = 4.5\,\mathrm{cm}^{-1}$ the pinned-mode resonance is clearly observed which becomes more pronounced as the temperatures is reduced below $T_{CDW} \approx 180\,\mathrm{K}$. The opening of the single particle gap causes the dielectric constant to increase drastically to approximately 700; the pinned-mode resonance leads to an additional contribution which is present already at room temperature due to fluctuations.

The strong influence of fluctuations in one dimension shifts the actual transition T_{CDW} well below the mean-field value $T_{MF} = \Delta/1.76k_B$. This intermediate temperature range $T_{CDW} < T < T_{MF}$ is characterized by the opening of a pseudogap in the density of states, i. e. a reduced intensity close to the Fermi energy which is observed in the magnetic susceptibility but not in dc transport. Optical experiments also see the development of the pseudogap and indications of the collective mode all the way up to T_{MF}. Utilizing a combination of different methods, the optical response of $K_{0.3}MoO_3$ was measured parallel and perpendicular to the highly conducting axis; the results for $T = 300\,\mathrm{K}$ and $200\,\mathrm{K}$ are displayed in Fig. 9. Clearly pronounced excitations are discovered in the spectra below $50\,\mathrm{cm}^{-1}$ for the electric field **E** parallel to the chains, the direction along which the charge-density wave develops below the Peierls transition temperature T_{CDW}. These excitations are associated with charge-density-wave fluctuations that exist even at room temperature and result in a collective contribution to the conductivity. A single optical experiment finally

brought a confirmation of this view: Fig. 10 exhibits results of transmission measurements through thin films of the blue bronze compound $Rb_{0.3}MoO_3$ on an Al_2O_3 substrate. The transmission coefficient and phase shift were recorded simultaneously using a Mach-Zehnder interferometer, which is driven by backward wave oscillators as powerful and tunable sources and which operates in the THz range of frequencies (30 GHz to 1500 THz, 1–50 cm^{-1}) [22].

3.2 Charge order

The crucial point of a CDW is the Fermi surface nesting; the driving force is the energy reduction of the occupied states right below the Fermi energy E_F when the superstructure is formed (cf. Fig. 7). Well distinct from a charge density wave is the occurrence of charge order (CO). The Coulomb repulsion V between adjacent lattice sites may lead to the preference of alternatingly more or less charge as depicted in Fig. 11. The extended Hubbard model is a good description of the relevant energies:

$$\mathcal{H} = -t \sum_{j=1} \sum_{\sigma=\uparrow\downarrow} \left(c^{+}_{j,\sigma} c_{j+1,\sigma} + c^{+}_{j+1,\sigma} c_{j,\sigma} \right) + U \sum_{j=1} n_{j\uparrow} n_{j\downarrow} + V \sum_{j=1} n_j n_{j+1} \,. \quad (1)$$

Here t denotes the hopping integral to describe the kinetic energy, U is the on-site Coulomb repulsion, and V is the nearest neighbour interaction. The disproportionation of charge on the molecules represents a short-range order and has to be commensurate with the lattice. CO may be accompanied by a slight lattice distortion (Fig. 11(d)), but this is a secondary effect. In contrast to a CDW, a metallic state above the ordering temperature is not required. If it is the case (metallic state), the gap in the density of states due to the superstructure also causes a metal-insulator transition. The most intriguing example of a charge-order driven metal-to-insulator transition was found in the two-dimensional organic conductor α-(BEDT-TTF)$_2$I$_3$, and this kept the community puzzled for almost twenty years. Below $T_{CO} = 135$ K, the dc and microwave conductivity (first measured in the group of H.-W. Helberg) drops many orders of magnitude (Fig. 12), but no modification in the lattice is observed [23]. Only recently it was understood that electronic correlations are responsible for this phase transition. Optical experiments (Raman and infrared) reveal

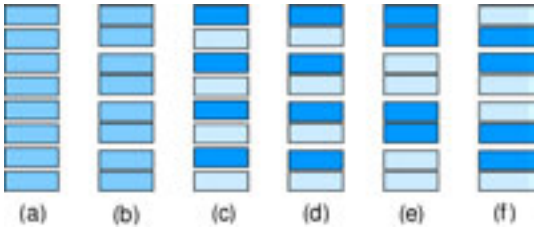


Figure 11. Possible arrangement of the molecules along the stacks. The disproportionation of charge is depicted by the different gray shade. The molecules can be dimerized (b), too, which may or may not be accompanied by charge order (c,d). The periodicity doubles again (teramerization) if neighbouring dimers carry different charge (e), but also if charge-rich molecules in adjacent dimers form pairs (f).

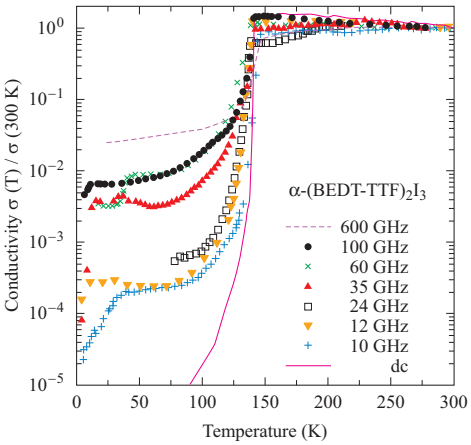


Figure 12. Temperature dependent conductivity of α-$(BEDT\text{-}TTF)_2I_3$ within the highly conducting plane measured by dc and microwave technique. The charge-order transition at 135 K leads to a rapid drop of the conductivity. The plateau in the conductivity between 40 K and 100 K increases with frequency indicating hopping conduction (after Ref. [23]).

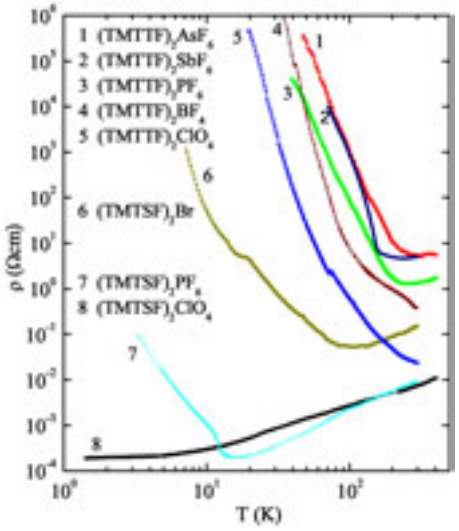


Figure 13. Temperature dependence of the dc resistivity of several Fabre and Bechgaard salts. As the temperature is reduced, the charges become increasingly localized in $(TMTSF)_2AsF_6$ and $(TMTSF)_2PF_6$, before the charge-ordered state is entered below 100 K. $(TMTSF)_2SbF_6$ shows a transition from a metal-like state directly into the charge-ordered state at $T_{CO} = 150$ K. $(TMTSF)_2PF_6$ undergoes a SDW transition at $T_{SDW} = 12$ K. Only $(TMTSF)_2ClO_4$ remains metallic all the way down to approximately 1.2 K where it becomes superconducting (after Ref. [25]).

a charge disproportionation from half a hole per molecule above the phase transition to $0.1e$ and $0.9e$ below T_{CO}; for a review see Dressel and Drichko [24].

Similar phenomena can also be observed in the quasi-one-dimensional $(TMTTF)_2X$ salts which are poor conductors at ambient temperature and exhibit a rapidly increasing resistivity as the temperature is lowered (Fig. 13). The reason is the accumulation of two effects which severely influence the energy bands as depicted in Fig. 14. The first one is a structural: due to the interaction with the anions (Fig. 5(c)) the molecular stack is dimerized as visualized in Fig. 11(b). The conduction band is split by a dimerization gap Δ_{dimer} and the material has a half-filled band. In a second step the Coulomb repulsion V causes charge disproportionation within the dimers (Fig. 11(d)). On-site Coulomb repulsion U also drives the one-dimensional half-filled system towards an insulating state: correlations induce a gap Δ_U at the Fermi energy E_F as shown in Fig. 14(c). The tetramerization of the CO according to Fig. 11(e) and f changes this picture conceptually (Fig. 14(d)): the soft gap Δ_{CO} due to short-range nearest-neighbour interaction V localizes the charge carriers. If not

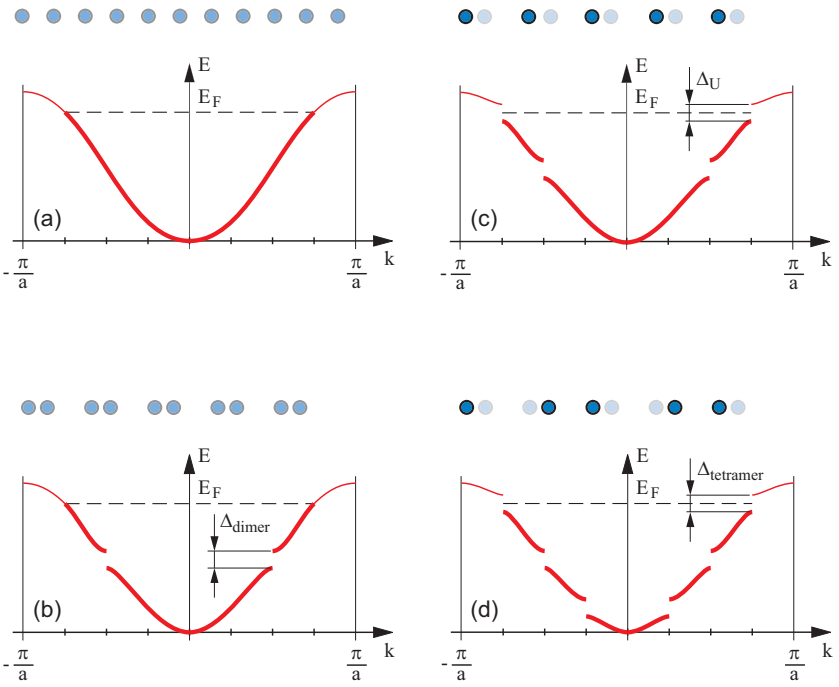


Figure 14. (a) A homogeneous stack of TMT*C*F, for example, with half an electronic charge $+e$ per molecule results in a three-quarter-filled band which leads to metallic behaviour. (b) Dimerization doubles the unit cell and the Brillouin zone is cut into two equal parts. The upper band is half filled and the physical properties remain basically unchanged. (c) Due to on-site Coulomb respulsion U a gap Δ_U opens at the Fermi energy E_F that drives a metal-to-insulator transition. (d) The tetramerization doubles the unit cell again and also causes a gap $\Delta_{tetramer}$.

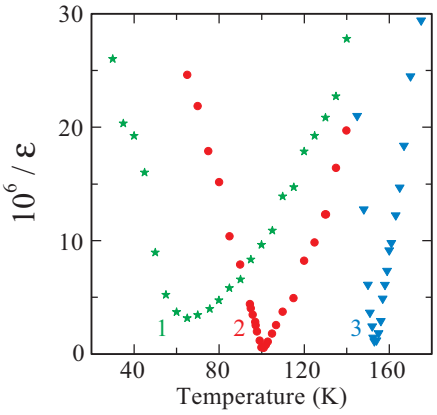


Figure 15. Temperature dependence of the inverse dielectric constant $1/\epsilon$ of $(TMTTF)_2X$, with different anions $X = PF_6$ (1), AsF_6 (2), and SbF_6 (3) (after Ref. [27]).

completely developed it just results in a reduction of the density of state (pseudogap). The tetramerization gap, on the other hand, is related to long-range order.

One- and two-dimensional NMR spectroscopy demonstrated the existence of an intermediate charge-ordered phase in the TMTTF family. At ambient temperature, the spectra are characteristic of nuclei in equivalent molecules. Below a continuous charge-ordering transition temperature T_{CO}, there is evidence for two inequivalent molecules with unequal electron densities. The absence of an associated magnetic anomaly indicates only the charge degrees of freedom are involved and the lack of evidence for a structural anomaly suggests that charge-lattice coupling is too weak to drive the transition [26].

The first indications of CO came from dielectric measurements in the radio-frequency range [27], where a divergency of the dielectric constant was observed at a certain temperature T_{CO}, as depicted in Fig. 15. Since this behaviour is well known from ferroelectric transitions, the idea is that at elevated temperatures the molecules carry equivalent charge of $+0.5e$; but upon lowering the temperature, the charge alternates by $\pm\rho$ causing a permanent dipole moment. Hence, new intermolecular vibrations at far-infrared frequencies below $100\,cm^{-1}$ get infrared active along all three crystal axes in the CO state due to the unequal charge distribution on the TMTTF molecules. Above the CO transition, these modes, which can be assigned to translational vibrations of the TMTTF molecules, are infrared silent but Raman active. By now there are no reports on a collective excitation which should show up as a low-frequency phonon.

The CO can be locally probed by intramolecular vibrations. Totally symmetric A_g modes are not infrared active; nevertheless, due to electron-molecular vibrational (emv) coupling (i. e. the charge transfer between two neighbouring organic TMTTF molecules which vibrate out-of phase), these modes can be observed by infrared spectroscopy for the polarization parallel to the stacks. As demonstrated in Fig. 16, the resonance frequency is a very sensitive measure of the charge per molecule [28]. The charge disproportionation increases as the temperature drops below T_{CO} in a mean-field fashion expected from a second-order transition; the ratio amounts to about 2:1 in $(TMTTF)_2AsF_6$ and 5:4 $(TMTTF)_2PF_6$. The charge disproportiona-

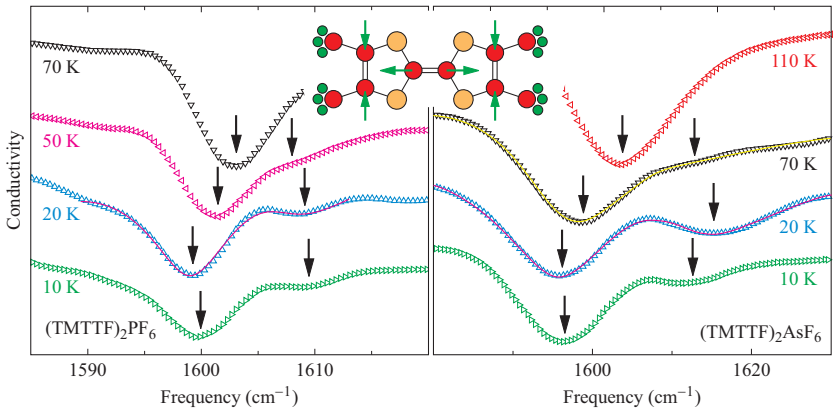


Figure 16. Mid-infrared conductivity of $(TMTTF)_2PF_6$ and $(TMTTF)_2AsF_6$ for light polarized parallel to the molecular stacks. The emv coupled totally symmetric intramolecular $\nu_3(A_g)$ mode (which mainly involves the C=C double bond) splits due to charge order as the temperature is cooled below T_{CO}. The charge disproportionation ratio amounts to about 2:1 in $(TMTTF)_2AsF_6$ and 5:4 $(TMTTF)_2PF_6$ (after Ref. [28]).

tion is slightly reduced in the AsF_6 salt, when it enters the spin-Peierls state, and unchanged in the antiferromagnetic PF_6 salt which infers the coexistence of charge order and spin-Peierls order at low temperatures.

3.3 Neutral-ionic transition

While in the previous example the crystals consist of separate cation and anion chains between which the electron transfer occurs, mixed-stack organic charge-transfer compounds have only one type of chain composed of alternating π electron donor and acceptor molecules (... $A^{-\rho}D^{+\rho}A^{-\rho}$ $D^{+\rho}A^{-\rho}D^{+\rho}$ ...) as sketched in Fig. 5(e). These materials are either neutral or ionic, but under the influence of pressure or temperature certain neutral compounds become ionic. There is a competition between the energy required for the formation of a D^+A^- pair and the Madelung energy. Neutral-ionic (NI) phase transitions are collective, one-dimensional charge-transfer phenomena occurring in mixed-stack charge-transfer crystals, and they are associated to many intriguing phenomena, as the dramatic increase in conductivity and dielectric constant at the transition, such as plotted in Fig. 17 [29,30].

In the simplest case the charge per molecule changes from completely neutral $\rho = 0$ to fully ionized $\rho = 1$. Ideally this redistribution of charge is decoupled from the lattice, and therefore should not change the inter-molecular spacing. In most real cases, however, the NI transition is characterized by the complex interplay between the average ionicity ρ on the molecular sites and the stack dimerization δ. The ionizity may act as an order parameter only in the case of discontinuous, first-order phase transitions. While the inter-site Coulomb interaction V favours a discontinuous jump of ionicity, the intra-chain charge-transfer integral t mixes the fully neutral and fully ionic quantum states and favours continuous changes in ρ.

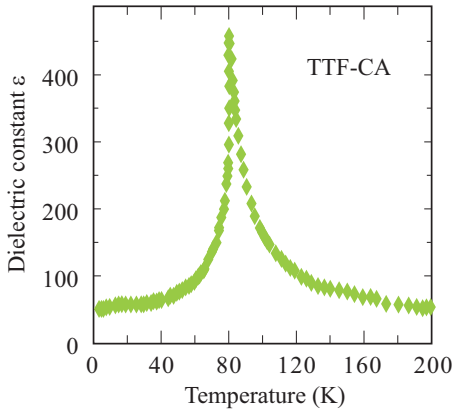


Figure 17. Temperature dependent dielectric constant $\epsilon(T)$ of TTF-CA measured at a frequency of 30 kHz (after Ref. [30]). The divergency at $T_{NI} = 81$ K clearly evidences the ferroelectric-like neutral-ionic transition.

The coupling of t to lattice phonons induces the dimerization of the stack, basically a Peierls-like transition to a ferroelectric state, which is a second-order phase transition. Intramolecular (Holstein) phonons, on the other hand, modulate the on-site energy U and favour a discontinuous jump in ρ.

In terms of a modified, one-dimensional Hubbard model [similar to Eq. (1)], the NI transition can be viewed as a transition from a band insulator to a Mott insulator, due to the competition between the energy difference between donor and acceptor sites, and the on-site Coulomb repulsion U. Peierls and Holstein phonons are both coupled to charge transfer electrons, albeit before the NI transition the former are only infrared active, and the latter only Raman active. This makes polarized Raman and reflection measurements a suitable tool to explore the NI transition.

The temperature induced NI transition of tetrathiafulvalene-chloranil (TTF-CA, cf. Fig. 5(d, e)) at $T_{NI} = 81$ K is the prime example of a first-order transition with a discontinuous jump in ρ. This can be seen in Fig. 18 by a jump in the frequency of those of the intramolecular vibrations, which are coupled to the electronic charge because their position depends on the charge on the molecules [31,32].

The vibronic bands present in the infrared spectra for $T > T_{NI}$ are combination modes involving the lattice mode, which gives rise to the Peierls distortion at the transition. From calculations we expect three lattice modes which couple to electrons and become stronger as the transition is approached. The lattice modes strongly couple to electrons and behave as soft modes of the ferroelectric transition at $T_{NI} = 81$ K. In Fig. 19 the low-frequency conductivity spectra are plotted for different temperatures $T > T_{NI}$. The lowest mode softens most and is seen strongly overdamped around $20\,\mathrm{cm}^{-1}$. The temperature evolution of this Peierls mode, which shows a clear softening (from 70 to $20\,\mathrm{cm}^{-1}$) before the first-order transition to the ionic ferroelectric state takes place. In the ordered phase, a clear identification and theoretical modelling of the Goldstone mode is still an open problem because the system has several degrees of freedom coupled to each other.

The cooperative charge transfer among the constructive molecules of TTF-CA can also be induced by irradiation by a short laser pulse. A photoinduced local charge-transfer excitation triggers the phase change and causes the transition in

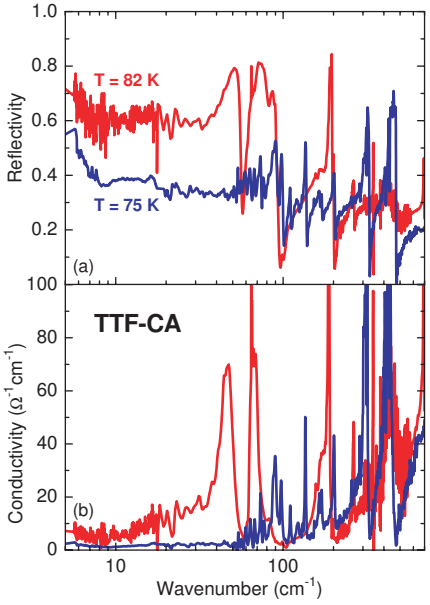


Figure 18. (a) Reflectivity and (b) conductivity spectra of TTF-CA measured along the stacking direction above (red line) and below (blue line) the neutral-ionic transition at $T_{\mathrm{NI}} = 81\,\mathrm{K}$ (after Ref. [32]).

both directions [33]. When Cl is replaced by Br in the tetrahalo-*p*-benzoquinones the lattice is expanded, (like a negative pressure) and the ionic phase vanishes completely. Hydrostatic pressure or Br-Cl substitution is utilized as a control parameter to tune the NI transition more or less continuously at $T \to 0$ [34].

4 Outlook

No doubt, one-dimensional physics matured from a toy model to an extremely active field of theoretical and experimental research, spanning a broad range from quantum gases to condensed-matter physics and semiconductor technology. Several novel and exciting phenomena can be investigated in these systems. In one-dimensional metals collective modes replace the single-particle excitations common to three-dimensional conductors and described by Landau's Fermi liquid concept of interaction electrons. Another property typical for low-dimensional solids is their susceptibility to symmetry breaking with respect to the lattice, the charge and the spin degrees of freedom. Broken-symmetry ground states imply that the system becomes stiff, because the modulation of the order parameter costs energy; therefore collective modes appear at low energies. In the case of magnets, the broken rotational symmetry leads to a magnetic stiffness and spin waves. In superconductors the gauge symmetry is broken, but due to the Higgs mechanism the Goldstone mode is absent at low frequencies and shifted well above the plasma frequency. In the examples above, we were dealing with translational symmetry, which is lowered in crystals due to charge ordering phenomena.

Charge density waves drive a metal to an insulator, for the Fermi surface becomes instable; the pinned-mode resonance can nicely be detected in the GHz range using

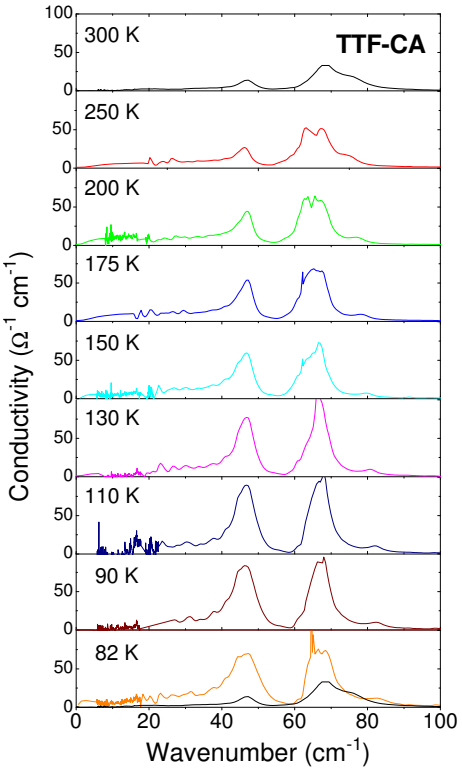


Figure 19. Low-frequency conductivity of TTF-CA for $T > T_{\mathrm{NI}}$ for different temperatures as indicated in the panels. As the NI transition is approached by decreasing temperature, the modes become stronger and an additional band appears as low as $20\,\mathrm{cm}^{-1}$. To make the comparison easier, the room temperature spectrum (black line) is replotted in the lowest frame (after Ref. [32]).

a variety of high-frequency and optical techniques. Purely electronic correlations between adjacent sites can cause charge disproportionation. Organic conductors are suitable realizations to investigate the properties at the metal-insulator transitions. The neutral-ionic transition observed in mixed-stack one-dimensional organic charge-transfer salts can be a pure change of ionizity, but commonly goes hand in hand with a Peierls distortion. This can be seen in a softening of the low-frequency phonon modes above the phase transition.

Optical methods and in particular microwave techniques as developed at the Dritte Physikalische Institut in Göttingen are powerful tools for investigation of charge-ordering phenomena in solids.

Acknowledgements. The review is based on many years of collaboration with a large number of people; only some of them can be mentioned here. In particular I would like to thank N. Drichko, M. Dumm, A. Girlando, B. Gorshunov, G. Grüner, and H.-W. Helberg.

References

[1] *Mathematical Physics in One Dimension*, edited by E. H. Lieb and D. C. Mattis (Academic Press, New York, 1966).

[2] T. Giamarchi, *Quantum Physics in One Dimension* (Oxford University Press, Oxford, 2004); M. Dressel, 'Spin-charge separation in quasi one-dimensional organic conductors', Naturwissenschaften **90**, 337 (2003).

[3] H. Moritz, T. Stöferle, M. Köhl, and T. Esslinger, 'Exciting collective oscillations in a trapped 1D gas', Phys. Rev. Lett. **91**, 250402 (2003).

[4] J. H. Davies, *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, Cambridge, 1998).

[5] F. J. Himpsel, A. Kirakosian, J. N. Crain, J.-L. Lin, und D. Y. Petrovykh, 'Self-assembly of one-dimensional nanostructures at silicon surfaces', Solid State Commun. **117**, 149 (2001).

[6] M. Dressel and H.-W. Helberg, 'AC conductivity of deformed germanium single crystals at $T = 4.2\,\mathrm{K}$', phys. stat. sol. (a) **96**, K199 (1986); M. Brohl, M. Dressel, H.-W. Helberg, and H. Alexander, 'Microwave conductivity investigations in plastically deformed silicon', Phil. Mag. B **61**, 97 (1990).

[7] H. Alexander and H. Teichler, 'Dislocations', in *Handbook of Semiconductor Technology*, Vol. 1, edited by K. A. Jackson and W. Schröter (Wiley-VCH, New York, 2000), p. 291.

[8] M. O'Connell, *Carbon Nanotubes* (Taylor & Francis, Boca Raton, 2006); P. J. F. Harris, *Carbon Nanotubes and Related Structures* (Cambridge University Press, Cambridge, 2004); S. Reich, C. Thomsen, and J. Maultzsch, *Carbon Nanotubes* (Wiley-VCH, Weinheim, 2004).

[9] P. Brüesch, 'Optical Properties of the One-Dimensional Pt Complex Compounds', in *One-Dimensional Conductors*, edited by H. G. Schuster (Springer-Verlag, Berlin, 1975), p. 194; P. Brüesch, S. Strässler, and H. R. Zeller, 'Fluctuations and order in a one-dimensional system. A spectroscopical study of the Peierls transition in $K_2Pt(CN)_4Br_{0.3}{\cdot}3(H_2O)$', Phys. Rev. B **12**, 219 (1975).

[10] P. Monceau (Ed.), *Electronic Properties of Inorganic Quasi-One-Dimensional Compounds*, Part I/II (Reidel, Dordrecht, 1985); C. Schlenker, J. Dumas, M. Greenblatt, and S. Van Smalen (Eds.), *Physics and Chemistry of Low Dimensional Inorganic Conductors* (Plenum, New York, 1996).

[11] P. A. Cox, *Transition Metal Oxides* (Clarendon Press, Oxford, 1992); S. Maekawa, T. Tohyama, S. E. Barnes, S. Ishihara, W. Koshibae, and G. Khaliullin, *The Physics of Transition Metal Oxides* (Springer, Berlin, 2004).

[12] D. Jérome and H. J. Schulz, 'Organic conductors and superconductors', Adv. Phys. **31**, 299 (1982); D. Jérome, in *Organic Conductors*, edited by J.-P. Farges (Marcel Dekker, New York, 1994), p. 405; M. Dressel, 'Spin-charge separation in quasi one-dimensional organic conductors', Naturwissenschaften **90**, 337 (2003); M. Dressel, 'Ordering phenomena in quasi one-dimensional organic conductors', Naturwissenschaften **94**, DOI 10.1007/s00114-007-0227-1 (2007).

[13] D. Jérome, A. Mazaud, M. Ribault, and K. Bechgaard, 'Superconductivity in a synthetic organic conductor $(TMTSF)_2PF_6$', J. Physique Lett. **41**, L95 (1980).

[14] J. Goldstone, 'Field theories with "superconductor" solution', Nuovo cimento **19**, 154 (1961); J. Goldstone, A. Salam, and S. Weinberg, 'Broken symmetries', Phys. Rev. **127**, 965 (1962).

[15] G. Grüner, *Density Waves in Solids* (Addison-Wesley, Reading, MA, 1994).

[16] M. Dressel and G. Grüner, *Electrodynamics of Solids* (Cambridge University Press, Cambridge, 2002).

[17] O. Klein, S. Donovan, M. Dressel, and G. Grüner, 'Microwave cavity perturbation technique. Part I: Principles', Int. J. Infrared and Millimeter Waves **14**, 2423 (1993); S. Donovan, O. Klein, M. Dressel, K. Holczer, and G. Grüner, 'Microwave cavity perturbation technique. Part II: Experimental scheme', Int. J. Infrared and Millimeter Waves **14**, 2459 (1993); M. Dressel, S. Donovan, O. Klein, and G. Grüner, 'Microwave cavity perturbation technique. Part III: Applications', Int. J. Infrared and Millimeter Waves **14**, 2489 (1993); A. Schwartz, M. Dressel, A. Blank, T. Csiba, G. Grüner, A. A Volkov, B. P. Gorshunov, and G. V. Kozlov, 'Resonant techniques for studying the complex electrodynamic response of conducting solids in the millimeter and submillimeter wave spectral range', Rev. Sci. Instrum. **66**, 2943 (1995); M. Dressel, O. Klein, S. Donovan, and G. Grüner, 'High frequency resonant techniques for the study of the complex electrodynamic response in solids', Ferroelectrics **176**, 285 (1996).

[18] H.-W. Helberg and B. Wartenberg, 'Zur Messung der Stoffkonstanten ϵ und μ im GHz-Bereich mit Resonatoren', Z. Angew. Phys. **20**, 505 (1966). The tradition goes back to the Institut für Angewandte Elektrizität (Institute of Applied Electricity) founded in the beginning of the 20th century and headed by Max Reich for a long time. Students like Arthur von Hippel spread this knowledge all around the world and made high-frequency investigations of solids to a powerful tool. The foundation of the *Laboratory of Insulation Research* and the *Radiation Laboratory* at MIT during World War II certainly had the largest impact.

[19] L. I. Buranov and I. F. Shchegolev, 'Method of measuring conductivity of small crystals at a frequency of 10^{10} Hz', Prib. Tekh. Eksp. (engl.) **14**, 528 (1971); I. F. Shchegolev, 'Electric and magnetic properties of linear conducting chains', phys. stat. sol. (a) **12**, 9 (1972); H.-W. Helberg and M. Dressel, 'Investigations of organic conductors by the Schegolev method', J. Phys. I. (France) **6**, 1683 (1996).

[20] B. P. Gorshunov, A. A Volkov, G. V. Kozlov, L. Degiorgi, A. Blank, T. Csiba, M. Dressel, Y. Kim, A. Schwartz, and G. Grüner, 'Charge density wave paraconductivity in $K_{0.3}MoO_3$', Phys. Rev. Lett. **73**, 308 (1994); A. Schwartz, M. Dressel, B. Alavi, A. Blank, S. Dubois, G. Grüner, B. P. Gorshunov, A. A. Volkov, G. V. Kozlov, S. Thieme, L. Degiorgi, and F. Lévy, 'Fluctuation effects on the electrodynamics of quasi one-dimensional conductors above the charge-density-wave transition', Phys. Rev. B **52**, 5643 (1995).

[21] A. V. Pronin, M. Dressel, A. Loidl, H. S. J. van der Zant, O. C. Mantel, and C. Dekker, 'Optical investiations of the collective transport in CDW-films', Physica B **244**, 103 (1998).

[22] G. Kozlov and A. Volkov, 'Coherent Source Submillimeter Wave Spectroscopy', in *Millimeter and Submillimeter Wave Spectroscopy of Solids*, edited by G. Grüner (Springer, Berlin, 1998), p. 51; B. Gorshunov, A. Volkov, I. Spektor, A. Prokhorov, A. Mukhin, M. Dressel, S. Uchida, and A. Loidl, 'Terahertz BWO-spectrosopy', Int. J. of Infrared and Millimeter Waves, **26**, 1217 (2005).

[23] K. Bender, K. Dietz, H. Endres, H.-W. Helberg, I. Hennig, H. J. Keller, H. W. Schäfer, and D. Schweitzer, '$(BEDT\text{-}TTF)^{2+}J_3^-$ - A two-dimensional organic metal', Mol. Cryst. Liq. Cryst. **107**, 45 (1984); M. Dressel, G. Grüner, J. P. Pouget, A. Breining, and D. Schweitzer, 'Field- and frequency dependent transport in the two-dimensional organic conductor α-$(BEDT\text{-}TTF)_2I_3$' J. de Phys. I (France) **4**, 579 (1994).

[24] M. Dressel and N. Drichko, 'Optical properties of two-dimensional organic conductors: signatures of charge ordering and correlation effects', Chem. Rev. **104**, 5689 (2004); M. Dressel, 'Ordering phenomena in quasi one-dimensional organic conductors', Naturwissenschaften **94**, DOI 10.1007/s00114-007-0227-1 (2007).

[25] M. Dressel, S. Kirchner, P. Hesse, G. Untereiner, M. Dumm, J. Hemberger, A. Loidl,

and L. Montgomery 'Spin and charge dynamics in Bechgaard salts', Synth. Met. **120**, 719 (2001).

[26] D. S. Chow, F. Zamborszky, B. Alavi, D. J. Tantillo, A. Baur, C. A. Merlic, and S. E. Brown, 'Charge ordering in the TMTTF family of molecular conductors', Phys. Rev. Lett. **85**, 1698 (2000).

[27] P. Monceau, F. Ya. Nad, and S. Brazovskii, 'Ferroelectric Mott-Hubbard phase of organic $(TMTTF)_2X$ conductors', Phys. Rev. Lett. **86**, 4080 (2001).

[28] M. Dumm, M. Abaker, and M. Dressel, 'Mid-infrared response of charge-ordered quasi-1D organic conductors $(TMTTF)_2X$', J. Phys. IV (France) **131**, 55 (2005).

[29] J. B. Torrance, J. E. Vazquez, J. J. Mayerle, and V. Y. Lee, 'Discovery of a neutral-to-ionic phase transition in organic materials', Phys. Rev. Lett. **46**, 253 (1981); J. B. Torrance, A. Girlando, J. J. Mayerle, J. I. Crowley, V. Y. Lee, P. Batail, and S. J. LaPlace, 'Anomalous nature of neutral-to-ionic phase transition in tetrathiafulvalene-chloranil', Phys. Rev. Lett. **47**, 1747 (1981); T. Mitani, Y. Kaneko, S. Tanuma, Y. Tokura, T. Koda, and G. Saito, 'Electric conductivity and phase diagram of a mixed-stack charge-transfer crystal: Tetrathiafulvalene-p-chloranil', Phys. Rev. B **35**, 427 (1987); S. Horiuchi, Y. Okimoto, R. Kumai, and Y. Tokura, 'Anomalous valence fluctuation near a ferroelectric transition in an organic charge-transfer complex', J. Phys. Soc. Japan **69**, 1302 (2000).

[30] S. Horiuchi, Y. Okimoto, R. Kumai, and Y. Tokura, 'Anomalous valence fluctuation near a ferroelectric transition in an organic charge-transfer complex', J. Phys. Soc. Jpn. **69**, 1302 (2000).

[31] M. Masino, A. Girlando, and Z. G. Soos, 'Evidence for a soft mode in the temperature induced neutral-ionic transition of TTF-CA', Chem. Phys. Lett. **369**, 428 (2003).

[32] M. Masino, A. Girlando, A. Brillante, R. G. Della Valle, E. Venuti, N. Drichko, and M. Dressel, 'Lattice dynamics of TTF-CA across the neutral ionic transition', Chem. Phys. **325**, 71 (2006); N. Drichko *et al.*, to be published.

[33] S. Y. Koshihara, Y. Takahashi, H. Saki, Y. Tokura, and T. Luty, 'Photoinduced cooperative charge transfer in low-dimensional organic crystals', J. Phys. Chem. B **103**, 2592 (1999).

[34] S. Horiuchi, Y. Okimoto, R. Kumai, and Y. Tokura, 'Quantum phase transition in organic charge-transfer complexes', Science **299**, 229 (2003).

Oscillations, Waves and Interactions, pp. 333–366
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-13-5

Multistep association of cations and anions. The Eigen-Tamm mechanism some decades later

Reinhard Pottel, Julian Haller, and Udo Kaatze

Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. Broadband ultrasonic absorption spectra and complex dielectric spectra for aqueous solutions of electrolytes are reported and are discussed in terms of cation-anion association schemes. Ultrasonic spectra for solutions of 3:2 valent salts clearly reveal the complete Eigen-Tamm multistep association mechanism, including inner-sphere, outer-sphere, and outer-outer-sphere complexes. Reduced association schemes follow for solutions of 2:2 valent and 2:1 valent salts, the latter just revealing the equilibrium between the complex of encounter and the outer-sphere complex. Dielectric and ultrasonic spectra are evaluated in terms of single steps in the intriguingly complex association scheme of $ZnCl_2$ aqueous solutions. The latter spectra are alternatively discussed assuming a fluctuating cluster model.

1 Introduction

The incomplete dissociation of multivalent salts in solutions is of considerable significance not just for the theory of electrolytes but also for biochemistry and wide fields of chemical engineering. The exploration of the molecular dynamics of electrolyte solutions and of the kinetics of ion complex formation has been an enduring topic of research at the Dritte Physikalische Institut from the very first. Interest was originally inspired by the technical problem of measuring distances in sea water by means of acoustical signals [1]. It was found that sea water may absorb sound more strongly than distilled water and that, in addition, the absorption depends in an unexpected manner upon the frequency of the sound field. Already in the early fifties of the last century Tamm and Kurtze developed techniques, enabling sound absorption measurements over the remarkably broad frequency range from 5 kHz to 300 MHz, and demonstrated the relaxation characteristics in the frequency dependent sonic absorption coefficient of 2:2 valent electrolyte solutions [2–7]. Consideration of sonic spectra also for solutions of 2:1 electrolytes lead to the conclusion that neither a simple inter-ionic interaction, without involvement of water, nor an interaction of cations or anions, respectively, just with water could be the reason for the ultrasonic excess absorption spectra of the aqueous systems [8]. As other effects, such as hydrolysis and ion cloud interactions were not consistent with the experimental findings, an interaction between cations, anions, and water molecules has been proposed the

cause for the absorption characteristics. This interaction involves a stepwise substitution of water molecules in the coordination shells of a cation-aquo complex by an anion [9,10]

$$\begin{aligned}(M^{m+})_{aq} + (L^{l-})_{aq} &\rightleftharpoons (M^{m+}(H_2O)_2L^{l-})_{aq} \\ &\rightleftharpoons (M^{m+}(H_2O)L^{l-})_{aq} \\ &\rightleftharpoons (ML)_{aq}^{(m-l)+}\end{aligned} \tag{1}$$

In this scheme $\left(M^{m+}\,(H_2O)_2\,L^{l-}\right)_{aq}$ denotes an outer-outer sphere complex in which the metal ion M^{m+} and the ligand L^{l-} are separated by two layers of water molecules. $\left(M^{m+}\,(H_2O)\,L^{l-}\right)_{aq}$ represents the outer sphere complex with one water layer between cation and anion, and $(ML)_{aq}^{(m-l)+}$ is the inner sphere complex, the contact ion pair.

Both former steps in the Eigen-Tamm scheme (Eq. (1)) contribute to the sonic excess absorption at high frequencies. In order to increase the significance of the spectra Plaß and Kehl extended the frequency range of measurements up to 2.8 GHz [11]. Nevertheless a clear conclusion on the number of relaxation terms within the experimental spectra was not reached and the existence of the second step in the complex formation scheme, reflecting the equilibrium between outer-outer sphere and outer sphere complexes, has been controversely discussed [12–16]. Combining frequency domain (ultrasonic) techniques and time domain (pressure-jump) techniques of the Drittes Physikalisches Institut and the Max-Planck-Institut für Biophysikalische Chemie, Göttingen, it was possible to clearly reveal three relaxation terms in the spectra of scandium sulfate solutions and to show thereby that all three steps in the reaction scheme (Eq. (1)) of the Eigen-Tamm mechanism may exist at least in 3:2 valent electrolyte solutions [17].

Complementary to the sonic spectrometry studies broadband dielectric measurement techniques have been developed and have been employed to verify the concept of ion complex formation. A detailed investigation into 2:2 valent electrolyte solutions was the first to identify dipolar ion structures by relaxation terms in the dielectric spectra of solutions [18,19]. Later 3:2 valent [20] as well as 1:1, 2:1, and 3:1 valent electrolytes [21] were studied. Particular attention was paid to zinc(II) chloride solutions with their intriguing complexation properties [22,23]. During the last years interest in dielectric relaxations due to ion-pairs in solutions has been rediscovered [24–29].

The studies of $ZnCl_2$ solutions have been supported by ultrasonic attenuation measurements which, in addition to any evaluations in terms of the conventional complexation scheme [30], have been alternatively dicussed employing a model of rapidly fluctuating ion clusters [31]. Recently, the Eigen-Tamm concept of stepwise association has been questioned even for 2:2 valent electrolytes and a mode-coupling theory has been presented assuming fluctuations in the ion concentration to be the cause of the high-frequency ultrasonic relaxation term of multivalent electrolytes instead of the formation of stoichiometrically well defined outer-sphere complexes [32].

Here we briefly review evidence in favour of the multistep association mechanism (Eq. (1)) of 2:2 and 3:2 valent electrolytes. We also discuss spectra of 1:1 and 2:1 valent salt solutions in which Coulombic interactions between cations and anions

are considerably reduced [30,31,33–35]. The latter include systems with calcium as cation, which have been investigated because of the far-reaching biochemical implications of that ion.

2 Experimental methods

2.1 Fundamental aspects

Sonic as well as dielectric spectrometry utilize naturally present molecular marks, the molar volume and the electrical dipole moment, respectively, to monitor the microdynamics and fast elementary kinetics of liquids (Fig. 1). Experimental techniques are currently available in the frequency range from 10^3 Hz to 10^{10} Hz for sonic absorption measurements [36,37] and in the even broader range from 10^{-6} Hz to 10^{12} Hz for the dielectric spectrometry [38]. Basically both methods aim at the study of the sample at thermal equilibrium. In order to reach sufficient accuracy in the measurements, however, in practice the sample is exposed to a small amplitude disturbing sonic or electromagnetic field, respectively. Methods in use consist in the observation of the response of the sample to either step pulses or harmonically alternating signals. The former method, globally named time-domain spectrometry uses pressure jumps, temperature jumps, or electrical field jumps to slightly disturb the system and to follow its relaxation into thermal equilibrium by continuous measurement of a suitable sample property, such as density, electrical conductivity, or dielectric polarization. Due to molecular interactions this property is unable to instantaneously obey the exterior force and therefore retardedly reaches its new equilibrium, typically following an exponential. From a fundamental point of view time domain techniques entail an unfavourable concentration of energy of the exciting signal in a short period of time. For this reason, frequency domain techniques, in which the sample response to a harmonically alternating acoustical or electromagnetic field is observed, are popular, particularly for high frequency measurements.

Because of the phase lag between the density and the pressure in the sonic wave and between the polarization and electrical field in the dielectric measurements, energy of the applied wave is absorbed in the liquid. The amplitude of a plane wave,

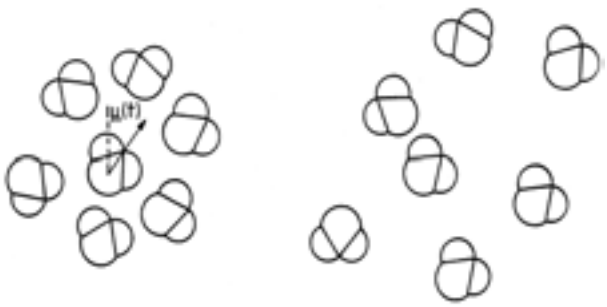


Figure 1. Ensemble of water molecules illustrating fluctuations in the direction of the electric dipole moment (left) and in the molar volume (right).

propagating within the sample, decreases exponentially along the axis of wave propagation. The absorption coefficient α in the exponential decay is one of the parameters that is determined in frequency domain spectrometry.

According to the Kramers-Kronig relations absorption is correlated with dispersion in the sound velocity and speed of light, respectively, within the sample. The dispersion in the sound velocity of liquids is notoriously small and thus normally not considered in the discussion of results. We mention, however, that the complexation kinetics of electrolyte solutions has been studied in the frequency range 3-200 MHz solely by high-precision sound velocity dispersion measurements [39,40].

In the dielectric spectrometry of dipolar liquids it is common practice to also measure dispersion and to verify consistency of the results thereby. Hence dielectric measurements typically involve complex quantities, preferably complex transfer functions or complex reflection coefficients of appropriate specimen cells, instead of only a scalar decay function. The dielectric properties of a sample are expressed in terms of a frequency dependent complex quantity, the permittivity

$$\epsilon(\nu) = \epsilon'(\nu) - i\epsilon''(\nu) = \frac{1}{\epsilon_0}\frac{P(t)}{E(t)} + 1\,. \tag{2}$$

Here $\epsilon'(\nu)$ and $\epsilon''(\nu)$ are the real part and negative imaginary part of the permittivity at frequency ν, $i^2 = -1$, and ϵ_0 is the electrical field constant. ϵ' represents the component of polarization P that is in phase with the electrical field $E = \hat{E} \cdot e^{i\omega t}$ and ϵ'' represents the contribution with a $\pi/2$ phase shift. The simplest relaxation spectral function is the Debye function [41]

$$\epsilon(\nu) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 + i\omega\tau} \tag{3}$$

with discrete relaxation time τ and angular frequency $\omega = 2\pi\nu$. This function corresponds with an exponential decay in the time domain. In the frequency range up to 100 GHz the dielectric spectrum of water can be well described by the Debye relaxation function [42,43]. As an example the water spectrum at 25°C is shown in Fig. 2 where also the extrapolated low-frequency ("static") permittivity $\epsilon(0)$ and the extrapolated high-frequency permittivity $\epsilon\,(\infty)$ are indicated.

Due to electrical conductivity ionic liquids, in which we are interested here, display an additional contribution

$$\epsilon''_\sigma(\nu) = \sigma/(\epsilon_0\omega) \tag{4}$$

to the total loss

$$\epsilon''_{tot}(\nu) = \epsilon''(\nu) + \epsilon''_\sigma(\nu)\,. \tag{5}$$

Because of the frequency dependence in ϵ''_σ, inversely proportional to ν, the conductivity contribution masks the dielectric contribution ϵ'' at low frequencies and renders measurements difficult or impossible at all at small ν. In Eq. (4) the specific electric conductivity σ is assumed independent of ν within the frequency range under consideration. Figure 3 illustrates the situation by a dielectric spectrum for an aqueous solution of sodium chloride. The salt concentration of that solution is less than one tenth of the salinity of the North Sea and is only one third of a physiological solution.

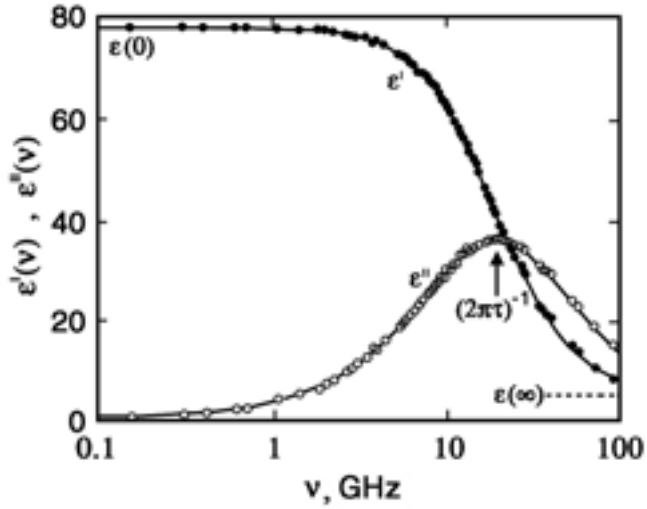


Figure 2. Real part $\epsilon'(\nu)$ and negative imaginary part $\epsilon''(\nu)$ of the complex dielectric spectrum of water at 25°C [43]. Permittivity data from the literature [44,45] are presented. Lines are graphs of Eq. (3) with parameter values from a regression analysis: $\epsilon(\infty) = 5.2, \epsilon(0) = 78.35$, and $\tau = 8.27$ ps [46].

A complementary situation exists in the ultrasonic spectrometry. Coupling of compressional waves to shear motion and, to lower extent, to heat conduction results in energy dissipation which manifests itself in the absorption coefficient by a contribution proportional to ν^2 if the shear viscosity and thermal conductivity themselves are independent of ν [49]. In the commonly used format, in which the absorption-per-wavelength, $\alpha\lambda$, is considered as a function of frequency, energy dissipation by shear viscosity and thermal conductivity add an asymptotic high-frequency term $B\nu$, with B independent of ν. An example for an ultrasonic spectrum is given in Fig. 4, where, along with the total absorption-per-wavelength, $\alpha\lambda$, the excess absorption is displayed,

$$(\alpha\lambda)_{exc} = \alpha\lambda - B\nu\,. \tag{6}$$

The $(\alpha\lambda)_{exc}$ data again follow a Debye-type relaxation function

$$(\alpha\lambda)_{exc} = \frac{A\omega\tau}{1+\omega^2\tau^2} \tag{7}$$

with discrete relaxation time τ and with amplitude A.

The background contribution restricts ultrasonic spectrometry at high frequencies. Additionally, the wavelength λ of the sonic field becomes so small (λ=150 nm at 10 GHz, water, 25°C) that, at even higher frequencies, the use of continuum models might be questioned.

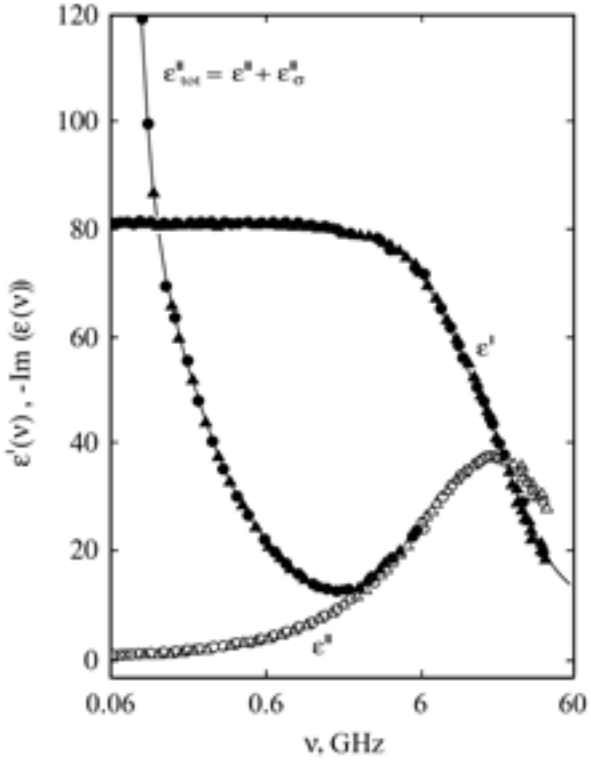


Figure 3. Real part $\epsilon'(\nu)$ and negative imaginary part $\epsilon''(\nu)$ of the complex dielectric spectrum of a solution of 0.051 mol/l NaCl in water at 20°C [47,48]. Open symbols represent the dielectric part $\epsilon'' = \epsilon''_{tot} - \epsilon''_{\sigma}$ in the total loss. Triangles on the one hand and points as well as circles on the other hand indicate results from two institutes [47].

2.2 Measurement techniques

Wide frequency ranges exist for both spectroscopic methods in which absolute measurements of the liquid properties are enabled by variation of the sample thickness. In dielectric measurements of aqueous systems the wavelength λ within the liquid at frequencies roughly above 1 GHz is sufficiently small to enable variable path length techniques for the determination of the complex propagation constant $\gamma = \alpha + i2\pi/\lambda$ [38]. In ultrasonic spectrometry the limiting parameter is the absorption coefficient which only above about 1 MHz is sufficiently large to enable absolute α measurements [36,37]. At lower frequencies quasistatic $\epsilon(\nu)$ measurements are performed with the aid of suitable sample cells, whereas the ultrasonic attenuation coefficient is obtained from resonator techniques in which the path length of interactions between the sample and the sonic field is virtually increased by multiple reflections.

Figure 5 shows the scheme of a semi-automatic double-beam interferometer for dielectric measurements in the microwave region [52]. It may be constructed from

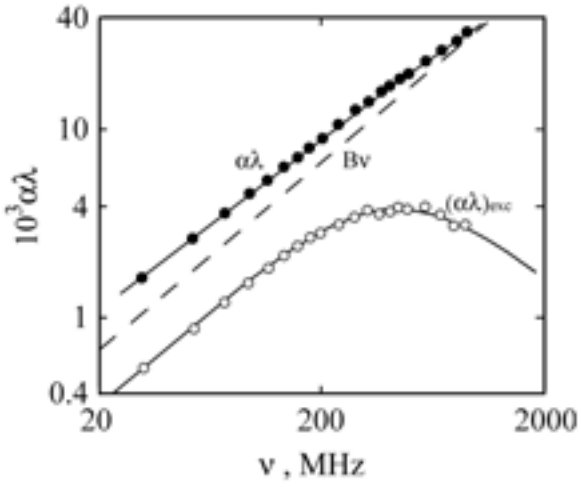


Figure 4. Spectra of the total ultrasonic absorption per wavelength (points), the asymptotic high frequency background contribution (dashed curve), and the excess absorption per wavelength (circles) for a 0.5 mol/l $CuCl_2$ solution in water at 25°C [50,51]. Lines are graphs of the analytical forms of $\alpha\lambda$ and $(\alpha\lambda)_{exc}$, using the relaxation spectral function defined by Eq.(7) and parameter values as following from a regression analysis of the experimental data: $A = 7.6 \cdot 10^{-3}, \tau = 0.35\,\text{ns}$, $B = 33.95\,\text{ps}$ [50].

coaxial-line components or waveguide devices. The central part is accentuated in grey. This part establishes a microwave bridge of which one branch contains the sample cell (6, Fig. 5), basically a circular waveguide or coaxial line, holding the liquid. Another waveguide or coaxial line (8) is immersed in the liquid. It can be precisely shifted along the direction z of wave propagation. The other branch is made of an attenuator (4) and a phase shifter (14) which, at a certain sample thickness z_0 can be adjusted to yield zero signal U_{out} at the interferometer output. Varying the position z of the probing waveguide or coaxial line and measuring U_{out} as the interferometer goes off-balance, along with the input signal U_{in}, yields the interferometer transfer function $T(z) = U_{out}(z)/U_{in}$. Fitting the theoretical $T(z)$ relation to the experimental data yields the propagation constant and thus the desired sample permittivity $\epsilon(\nu)$.

The corresponding set-up for the ultrasonic spectrometry is sketched in Fig. 6. Pulse-modulated waves are applied in order to allow for the separation of the measurement signal from electrical crosstalk and also from waveforms resulting from multiple reflections [53,54]. Therefore, instead of an interferometer, a comparator set-up is employed, in which coaxial switches (5, Fig. 6) allow the measurement branch, containing the sample cell (8), to be replaced by a high presicion piston attenuator (13), featuring a calculable signal vernier [55]. The scalar transfer function $T(z)$ of the sample cell at varying sample thickness z allows for a straightforward evaluation in terms of the absorption coefficient of the liquid. From a standing-wave pattern in

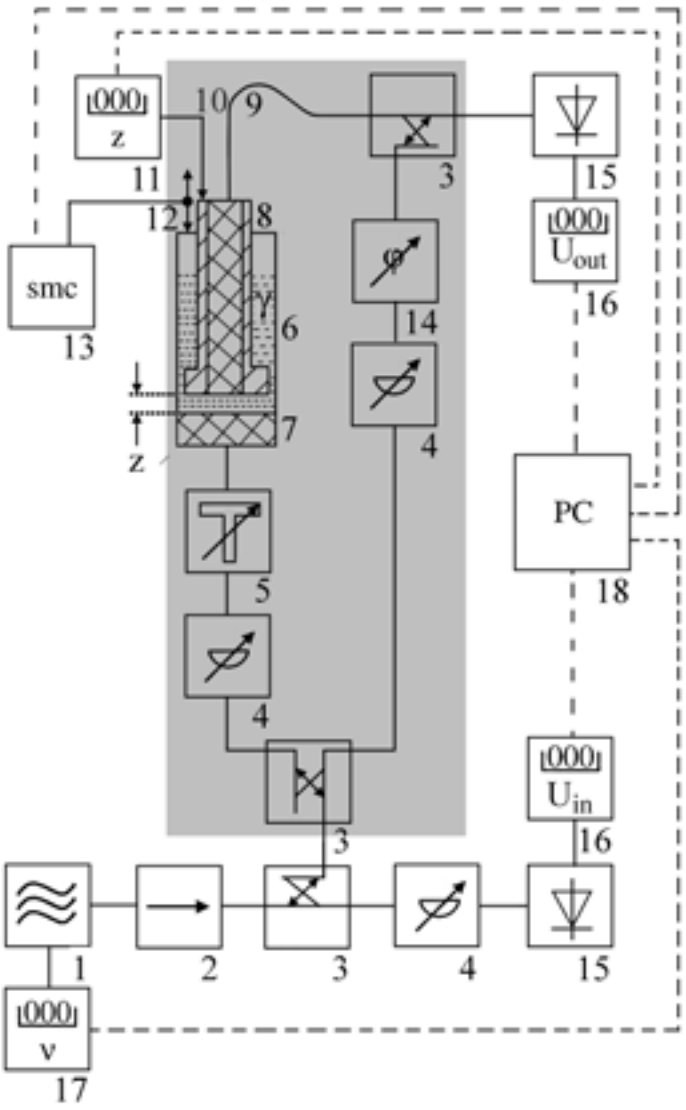


Figure 5. Block diagram of semi-automatic microwave double-beam interferometers for the dielectric spectrometry of liquids [52]: 1, signal generator; 2, uniline; 3, directional coupler; 4, variable attenuator; 5, impedance transformer; 6, sample cell with 7, sealing dielectric window and 8, precisely shiftable probe; 9, flexible waveguide or coaxial line; 10, digital distance meter with 11, control unit; 12, stepping motor with 13, control unit; 14, phase shifter; 15, power sensor; 16, level meter; 17, frequency counter; 18, process control computer.

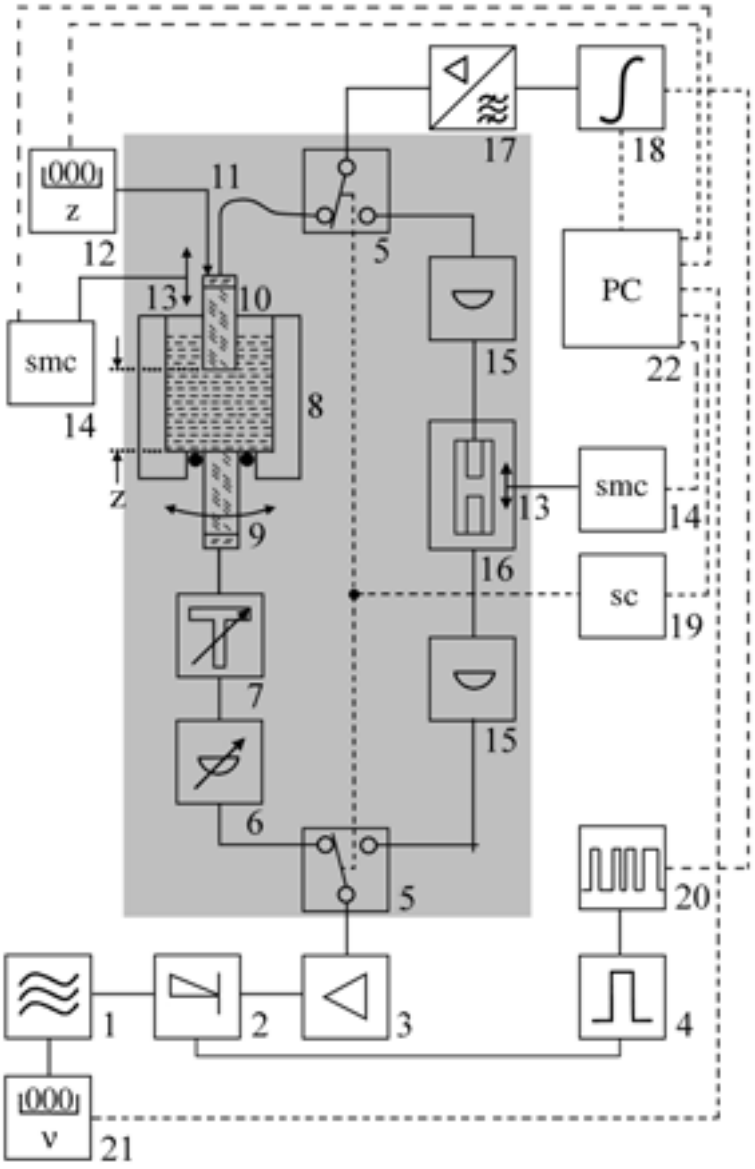


Figure 6. Block diagram of the semi-automatic comparator set-up for absolute ultrasonic absorption coefficient measurements between 1 MHz and 5 GHz [53,54]: 1, signal generator; 2, modulator; 3, broadband power amplifier; 4, pulse generator; 5, change-over switch; 6, variable attenuator; 7, impedance transformer; 8, specimen cell with 9, transmitter unit prepared for parallel adjustment to 10, receiver unit; 11, digital distance meter with 12, control unit; 13, stepping motor drive or piezo-translator with 14, control unit; 15, fixed coaxial attenuator for impedance matching; 16, high-precision adjustable piston attenuator; 17, superheterodyne receiver; 18, boxcar integrator; 19, switch driver and control; 20, trigger and control-pulse generator; 21, frequency counter; 22, process control computer.

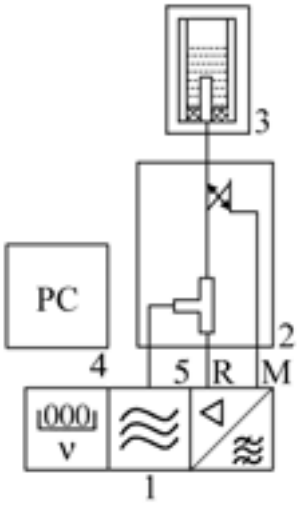


Figure 7. Principle of quasistatic electrical input impedance measurement system [56]: 1, network analyzer with S, signal output port, as well as R, reference and M, measurement signal input port; 2, reflection test set; 3, sample cell; 4, process control computer.

$T(z)$ at small sample thickness z also the sonic wavelength λ within the sample can be derived.

In dielectric spectrometry at frequencies below 3 GHz quasistatic electric input impedance measurements can be performed utilizing a suitable network analyzer with reflection test set (Fig. 7). Sample cells from the cut-off variety [56] have proven well suited. Such cells basically consist of a coaxial line/circular waveguide transition, the waveguide diameter being sufficiently small to excite the device, filled with the sample liquid, below the cutoff frequency of its fundamental TM_{01} field mode. The evanescent electromagnetic field in the waveguide section, the field in the feeding coaxial line, as well as the field lines passing the sealing dielectric window can be represented by a network of capacitors, the capacitances of which are determined by calibration measurements [46,56,57].

Because of the frequency dependence of the $B\nu$-term in the total sonic absorption per wavelength (Eq. 6) the sonic absorption coefficient of aqueous solutions is small at frequencies below about 10 MHz. To increase the sensitivity in the ultrasonic absorption spectrometry at low frequencies, resonator methods are employed. The superior measuring ability of network analyzers is again utilized. It is convenient, however, to determine the transfer function of the resonator rather than its reflection coefficient (Fig. 8). The liquid is contained in a circular cylindrical cavity. The end faces of the cell are normally formed by piezoelectric quartz transducers [58–60], but reflectors made of glass, with transducers attached to their back faces, are also applied [61]. In order to reduce diffraction of the sonic waves at the low-frequency measurement

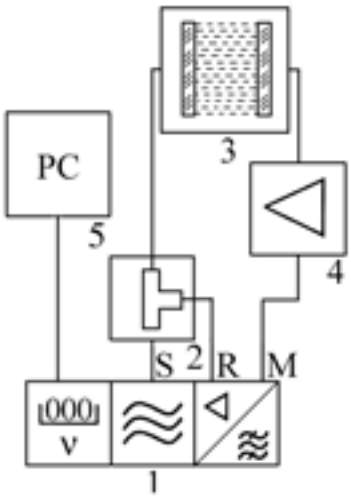


Figure 8. Schematic representation of sonic resonator measurement set-up [58–61]: 1, network analyzer as in Fig. 7; 2, signal divider; 3, cavity resonator cell; 4, amplifier; 5, process control computer.

range of the device, concavely shaped end faces, with radius of curvature at around 2 m are employed [59–61]. Using the resonator method, the absorption coefficient α of the sample liquid is determined relative to a reference liquid with matched sound velocity and density [62].

3 Aqueous solutions of 2:2 and 3:2 valent salts

3.1 Ultrasonic spectra

As an example for 2:2 electrolytes an ultrasonic excess absorption spectrum for a solution of 0.1 mol/l $MnSO_4$ in water is displayed in Fig. 9. The line additionally given in that diagram represents the sum of two Debye relaxation terms

$$(\alpha\lambda)_{exc} = \frac{A_1\omega\tau_1}{1+\omega^2\tau_1^2} + \frac{A_2\omega\tau_2}{1+\omega^2\tau_2^2}\,. \tag{8}$$

Within the broad frequency range of measurements the two-Debye-term model represents the measured data within the limits of eperimental errors. Hence the spectrum likely reflects two steps of the Eigen-Tamm association scheme (Eq. (1)), either

$$\left(Mn^{2+}\,(H_2O)_2\,SO_4^{2-}\right)_{aq} \rightleftharpoons \left(Mn^{2+}\,(H_2O)\,SO_4^{2-}\right)_{aq} \rightleftharpoons \left(Mn^{2+}SO_4^{2-}\right)_{aq} \tag{9}$$

or

$$\left(Mn^{2+}\right)_{aq} + \left(SO_4^{2-}\right)_{aq} \rightleftharpoons \left(Mn^{2+}\,(H_2O)\,SO_4^{2-}\right)_{aq} \rightleftharpoons \left(Mn^{2+}SO_4^{2-}\right)_{aq}\,. \tag{10}$$

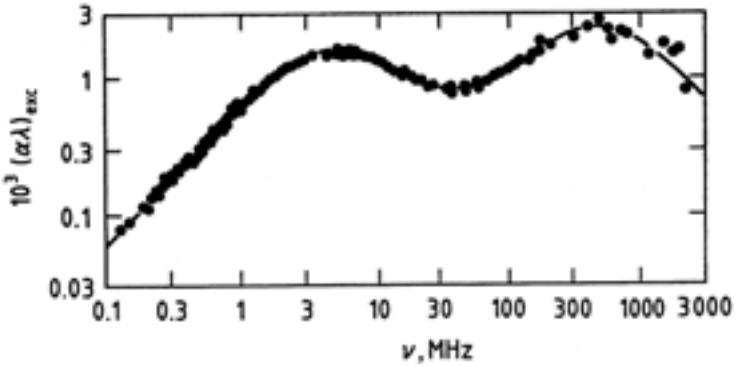


Figure 9. Ultrasonic excess absorption spectrum for a solution of 0.1 mol/l $MnSO_4$ in water at 25°C. The data are taken from the literature [58,60,65], the line is the graph of a superposition of two Debye type relaxation terms (Eq. (8)).

The former scheme (Eq. (9)) suggests another relaxation term to contribute to the spectrum at frequencies above the measuring range, namely the term due to the equilibrium between the completely dissociated ions and the outer-outer-sphere complexes. The latter scheme (Eq. (10)) is based on the assumption of a negligibly small concentration of outer-outer-sphere complexes.

In the early ultrasonic relaxation studies of ion complex formation [5–11,15] the excess absorption of the electrolyte solutions has been calculated using the B value of water in Eq. (6). The high frequency relaxation region was found to extend over a broader frequency band than a single Debye term. For this reason the excess absorption spectra were discussed in terms of three Debye-type relaxation processes which were assigned to the three steps in the complete Eigen-Tamm scheme (Eq. (1)). This assumption has been questioned [12,16,63] because of the unphysical volume change for the formation of outer-sphere complexes from outer-outer-sphere complexes following thereby. An unusually large change in volume results which is furthermore negative. Since B depends on the shear viscosity η_S of the liquid and as η_S will change on addition of salt, use of the B value of the solvent in the evaluation of the electrolyte solution spectra is doubtless an approximation. Due to the dominance of the background absorption at high frequencies, however, small errors in B may virtually broaden the relaxation characteristics of that part of the ultrasonic spectra. Such broadening of relaxation regions is illustrated in Fig. 10. Hence the assumption of all three steps in the Eigen-Tamm scheme to show up in the ultrasonic spectra of 2:2 valent salt solutions seems to rely on a slightly incorrect background absorption. Evidence for the existence of the outer-outer sphere complex may be obtained from consistency of relaxation parameters, measured as a function of salt concentration, with the kinetic relations [64].

Recently an alternative model for the description of ultrasonic spectra has been proposed [63] in which the low frequency Debye-type relaxation in the spectra of 2:2 valent electrolyte solutions is furtheron assigned to the formation of inner-sphere

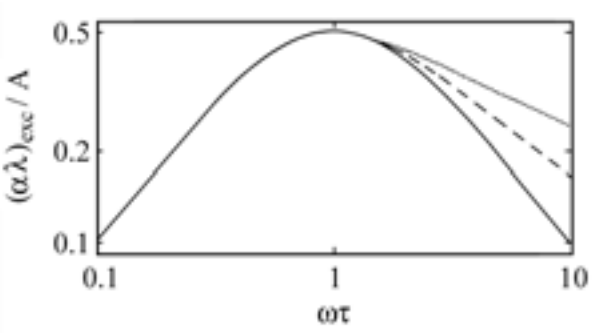


Figure 10. Graph of a Debye relaxation function $(\alpha\lambda)_{exc} = A\omega\tau/(1+\omega^2\tau^2)$, full line and of a Debye function with small differential term $\Delta B\nu$ added: $\Delta B = 0.02\pi\tau A$, dashed line; $\Delta B = 0.04\pi\tau A$, dotted line.

complexes. The high-frequency relaxation region, however, is discussed in terms of a distribution function theory in which no distinct ion complexes are assumed a priori [32]. Another approach that does not proceed from the existence of stoichiometrically defined complex structures but from long-range concentration fluctuations is briefly presented in the discussion of zinc chloride solution spectra.

Figure 11 shows ultrasonic excess attenuation spectra for solutions of scandium sulfate in water. These spectra obviously reveal three relaxation regions within the frequency range of measurements. The three Debye relaxation terms are indicated by dashed curves. They have been assigned to the coupled scheme [17]

$$\begin{aligned}\left(\mathrm{Sc}^{3+}\right)_{aq} + \left(\mathrm{SO}_4^{2-}\right)_{aq} &\underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \left(\mathrm{Sc}^{3+}\,(\mathrm{H_2O})_2\,\mathrm{SO}_4^{2-}\right)_{aq} \\ &\underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} \left(\mathrm{Sc}^{3+}\,(\mathrm{H_2O})\,\mathrm{SO}_4^{2-}\right)_{aq} \underset{k_{43}}{\overset{k_{34}}{\rightleftharpoons}} \left(\mathrm{ScSO_4}\right)^{+}_{aq}\,,\end{aligned} \tag{11}$$

corresponding with Eq. (1). Evaluation of the relaxation parameters of $Sc_2(SO_4)_3$-solution spectra at salt concentrations between 0.0033 and 0.1 mol/l enabled a complete characterization of the stepwise association scheme. As the pH of the solutions was adjusted at 2.4, at which the sulfate ion is partly protonated, the protolysis reaction

$$\left(\mathrm{Sc}^{3+}\right)_{aq} + \left(\mathrm{SO}_4^{2-}\right)_{aq} + \mathrm{H_3O}^{+} \rightleftharpoons \left(\mathrm{Sc}^{3+}\right)_{\mathrm{aq}} + \left(\mathrm{HSO}_4^{-}\right)_{\mathrm{aq}} + \mathrm{H_2O}\,, \tag{12}$$

coupled to the stepwise association mechanism, has been also taken into account [17]. The forward and reverse rate constants k_{ij} and k_{ji}, respectively, following from the evaluation of the ultrasonic spectra are listed in Table 1 where also the equilibrium constants

$$K_i = k_{ij}/k_{ji}\,, \tag{13}$$

the association constant

$$K_a = K_1[1 + K_2(1 + K_3)] \tag{14}$$

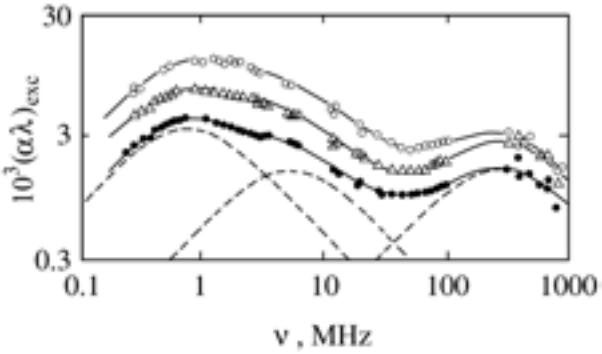


Figure 11. Ultrasonic excess absorption spectra for solutions of $Sc_2(SO_4)_3$ in water at 25°C and at pH=2.4 (○, 0.1 mol/l; △, 0.058 mol/l; •, 0.033 mol/l [17]). Dashed lines show the subdivision of the latter spectrum into three Debye-type relaxation terms as following from a nonlinear least-squares regression analysis of the experimental data. Full lines are graphs of the sum of these terms with the parameter values found by the fitting procedure.

and the reaction volumes ΔV_i are given. The values obtained solely from ultrasonic measurements are reasonable and support the assumption of the Eigen-Tamm multistep association mechanism.

3.2 Dielectric spectra

The complex dielectric spectrum for a solution of 0.1 mol/l $Al_2(SO_4)_3$ in water differs from that of the solvent in various aspects (Fig. 12). Both spectra display a dispersion $(d\epsilon'(\nu)/d\nu < 0)$/absorption $(\epsilon''(\nu) > 0)$ region in the frequency range around 20 GHz reflecting the dielectric relaxation of water. The solvent contribution ϵ_1 to the extrapolated static permittivity $\epsilon(0)$ of the solution, however, is considerably smaller than the static permittivity $\epsilon_W(0)$ of water. Also the relaxation frequency $(2\pi\tau_1)^{-1}$ of the water contribution to the solution spectrum is slightly shifted with respect to the relaxation frequency $(2\pi\tau_W)^{-1}$ of water at the same temperature. Both changes

i,j	k_{ij} s^{-1}	k_{ji} s^{-1}	K_i	ΔV_i cm^3/mol
1,2	$(2 \pm 0.5) \cdot 10^{11}$ §	$(9 \pm 3) \cdot 10^8$	220 ± 100 *	13 ± 3
2,3	$(1.4 \pm 0.5) \cdot 10^7$	$(2 \pm 0.7) \cdot 10^7$	0.7 ± 0.5	7 ± 2
3,4	$(9.9 \pm 3) \cdot 10^6$	$(3.3 \pm 1) \cdot 10^6$	3.0 ± 0.7	7 ± 2
		$K_a = (800 \pm 400)(mol/l)^{-1}$		

Table 1. Rate constants k_{ij} and k_{ji} of the coupled reaction scheme of ion association (Eq.11), equilibrium constants K_i and reaction volumes ΔV_i, as well as association constant K_a (Eq.14) for aqueous solutions of scandium sulfate [17]; § $s^{-1}(mol/l)^{-1}$, *$(mol/l)^{-1}$.

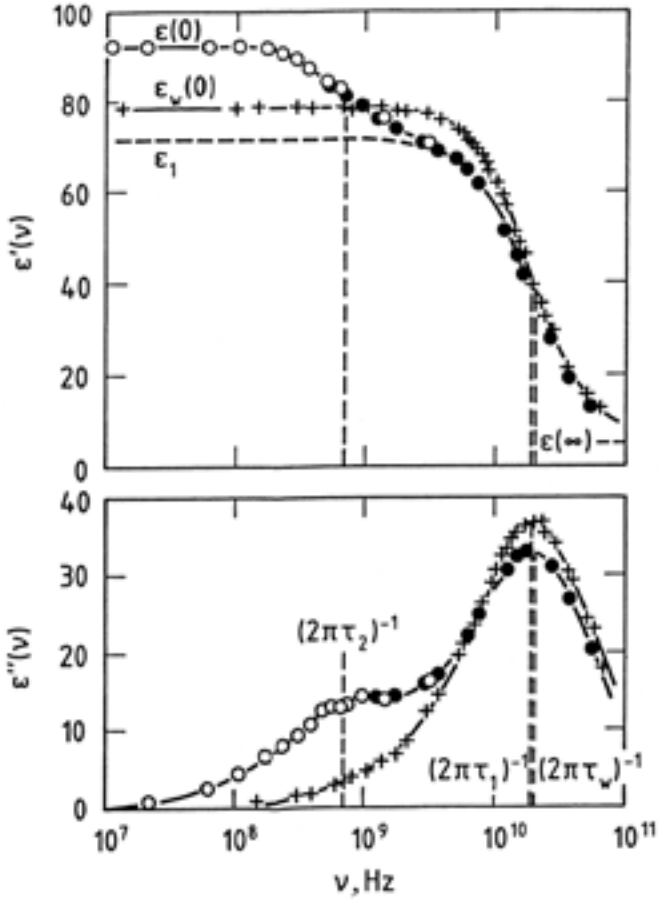


Figure 12. Real part $\epsilon'(\nu)$ and negative imaginary part $\epsilon''(\nu)$ of the complex dielectric spectrum of water (+) and of a 0.1 mol/l solution of $Al_2(SO_4)_3$ in water ($\circ, \bullet$) at 25°C [20]. Open symbols indicate data from a difference method applied in time domain measurements, closed symbols show data from frequency domain techniques. Lines are graphs of the Debye function (Eq. (3)) with subscript "w" refering to water and of the two-Cole-Cole-term function defined by Eq. (15).

in the solvent contribution to the complex dielectric spectrum are due to interactions of the water molecules with the ionic species. A remarkable feature in the spectrum of the electrolyte solution (Fig. 12) is the additional low-frequency relaxation with amplitude $\epsilon(0) - \epsilon_1$ and relaxation frequency $(2\pi\tau_2)^{-1}$. This relaxation reveals directly the existence of dipolar ion complex structures with life times larger than the reorientation times.

The relaxations in the solution spectrum are subject to a small distribution of relaxation times. The spectrum, without conductivity contribution, has been repre-

sented assuming a sum of Cole-Cole terms [66]

$$\epsilon(\nu) = \epsilon(\infty) + \frac{\epsilon_1 - \epsilon(\infty)}{1 + (i\omega\tau_1)^{(1-h_1)}} + \frac{\epsilon(0) - \epsilon_1}{1 + (i\omega\tau_2)^{(1-h_2)}}\,. \tag{15}$$

These terms reflect a relaxation time distribution function $G(\tau)$ which, when $\tau G(\tau)$ is plotted versus $\ln(\tau/\tau_\mathrm{i})$, $i = 1, 2$, is symmetrically bell shaped around $\tau/\tau_\mathrm{i} = 1$. Here τ_1 and τ_2 denote the principal relaxation times and parameters h_1 and h_2, $0 \leq h_1, h_2 < 1$ measure the width of the distribution function.

Despite of the clear indications of ion complex structures by the low-frequency relaxation term in the dielectric spectra the finding of only one relaxation regime for the ion processes is, on a first glance, a surprising result. It is common to all dielectric studies of ion complex formation [18–29,67–70]. Absence of distinct relaxation terms for the different ion species in the dielectric spectra becomes particularly evident when comparison to the corresponding ultrasonic spectra is made (Fig. 13). Interesting, the relaxation frequency $(2\pi\tau_2)^{-1}$ in the dielectric spectrum shown in Fig. 13 is larger than the largest relaxation frequency in the ultrasonic excess absorption spectrum. Obviously, in the dielectric spectra of the electrolyte solutions the complex formation/decay processes are short-circuited by the faster reoriental motions.

Based on the Debye model of rotational diffusion [41] reorientation times for the outer-outer-sphere, the outer-sphere, as well as the inner-sphere complexes have been estimated which are too close to each other to allow for a clear separation of the

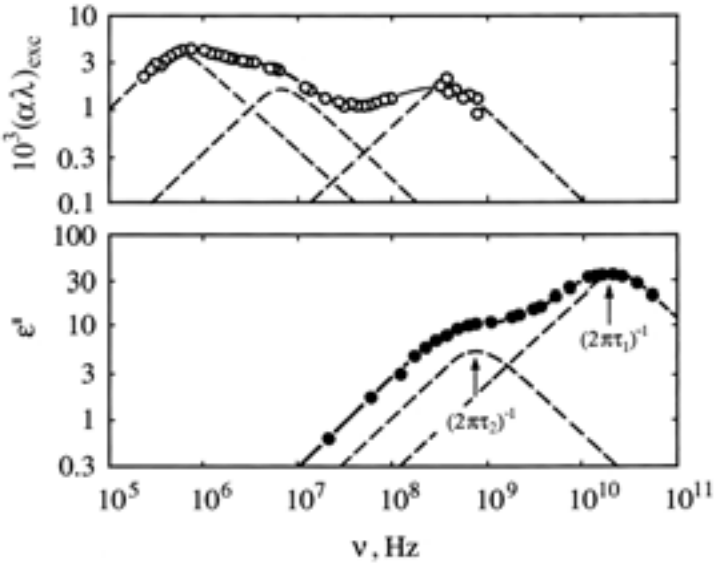


Figure 13. Ultrasonic excess absorption (○) and dielectric loss spectrum (●) for the solution of 0.1 mol/l $Sc_2(SO_4)_3$ in water at 25°C [20].

broadband relaxation contributions from the different ion complex species [20]. Dielectric spectroscopy thus yields direct evidence for the existence of ion pairs with life times exceeding the reorientation times, respectively, but a deconvolution of the experimental spectra is normally only possible when reasonable assumptions on the underlying molecular processes are made.

4 Electrolyte solutions of monovalent anions

4.1 Evidence from the solvent contribution to the static permittivity

As revealed by the dielectric spectrum of the aluminium sulfate solution displayed in Fig. 12 the solvent contribution ϵ_1 to the extrapolated static permittivity may fall significantly below the solvent permittivity $\epsilon_v\,(0)$ $(= \epsilon_w\,(0))$. The reduction in ϵ_1 is partly due to the dilution of the dipolar solvent by the solute. This part in the polarization deficiency can be considered by a suitable mixture relation, e.g. the Bruggeman formula [71]

$$\left(\frac{\epsilon_v(0)}{\epsilon(0)}\right)^{1/3} = (1-v)\frac{\epsilon_v(0)-\epsilon_u}{\epsilon(0)-\epsilon_u} \tag{16}$$

for the resulting permittivity $\epsilon\,(0)$ of a solution of spherical particles with volume fraction v and frequency independent permittivity ϵ_u in solvent with permittivity $\epsilon_v\,(0)$. In addition, two other effects may contribute to the reduction in ϵ_1. One effect is suggested by extrapolated static permittivity data as shown in Fig. 14.

The dielectric spectra of the bromide salt solutions for which the $\epsilon\,(0)$ data are presented, within the frequency range of measurements, do not indicate contributions from ion complexes, thus $\epsilon_1 = \epsilon\,(0)$ with these systems. For the bromides of large organic cations the $\epsilon\,(0)$ values slightly exceed the predictions of the Bruggeman mixture relation (Eq. (16)). This tendency in the extrapolated static permittivity seems to be characteristic to aqueous solutions of organic solutes and is assumed to be due to hydrophobic interaction effects [74–76]. The $\epsilon\,(0)$ value of the alkali halide solutions are smaller than predicted by Eq. (16) and, furtheron, the deviation from the mixture relation increases with decreasing cation radius. This feature points at an interaction between the dipole moment of the solvent molecules and the electric field of the small cations. The preferential orientation of the dipole moments within the Coulombic fields (Fig. 15) leads to a reduced orientation polarizability of the solvent, usually named "dielectric saturation" [72,77]. We shall get back to saturation effects later.

The other effect that leads to a reduction in the extrapolated static permittivity of electrolyte solutions is featured by the $\epsilon\,(0)$ data for solutions of lithium chloride in two different solvents, given in Fig. 16. For methanol solutions the reduction in the permittivity ratio $\epsilon\,(0)\,/\epsilon_v\,(0)$ is considerably larger than for aqueous solutions. These findings reflect a feature of the kinetic polarization deficiency [80–84] resulting from a coupling of dielectric properties to the hydrodynamics of the conducting liquids. An ion moving in a liquid, that is exposed to an external electric field, sets up a nonuniform flow in its ambient solvent [85]. The dipole moments of the solvent

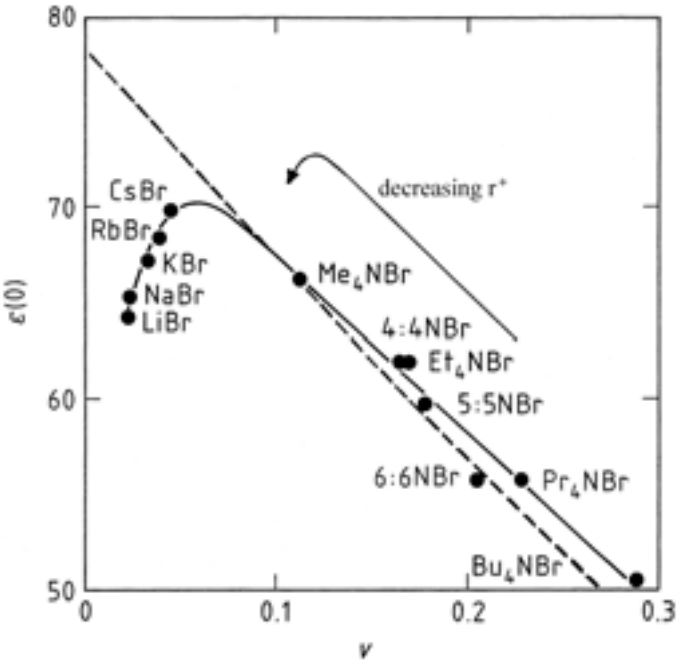


Figure 14. Extrapolated static permittivity $\epsilon(0)$ versus volume fraction v of solute for 1 mol/l aqueous solutions of alkali bromides [72] as well as tetraalkylammonium and azoniaspiroalkane bromides at 25°C: Me_4NBr, tetramethylammonium bromide; Et_4NBr, tetraethylammonium bromide; Pr_4NBr, tetrapropylammonium bromide; Bu_4NBr, tetrabutylammonium bromide; 4:4NBr, 5-azoniaspiro[4,4]nonane bromide; 5:5NBr, 6-azoniaspiro[5,5]undecane bromide; 6:6NBr, 7-azoniaspiro[6,6]tridecane bromide [73]. The dashed line is the graph of the Bruggeman mixture relation (Eq. (16)). The full line is drawn to guide the eyes.

molecules are turned thereby in the direction opposed to the one given by the external field. In the Hubbard-Onsager continuum theory a dielectric decrement

$$\delta\epsilon_{HO} = \frac{2}{3}\frac{\epsilon_v(0) - \epsilon_v(\infty)}{\epsilon_0\epsilon_v(0)}\sigma\tau_v \tag{17}$$

follows, where $\epsilon_v(\infty)$ and τ_v denote the extrapolated high-frequency permittivity and dielectric relaxation time of the solvent, respectively. In deriving Eq. (17) perfect slip boundary conditions on the solvent flow at the ion surfaces have been assumed. The large differences between the $\epsilon(0)$ data for the two series of LiCl solutions in Fig. 16 obviously reflect the different relaxation times of the solvent ($\tau_v = (8.27 \pm 0.05)$ ps, water [46]; $\tau_v = (48.7 \pm 1)$ ps, methanol [46]; 25°C).

Equation (17) allows the kinetic depolarization decrement to be calculated and, considering also the small reduction in the $\epsilon(0)$ values from the dilution of the dipolar solvent (Eq. (16)), the remaining polarization deficiency to be evaluated in terms of dielectric saturation. The degree of saturation is expressed by numbers $Z^{\pm}$ of

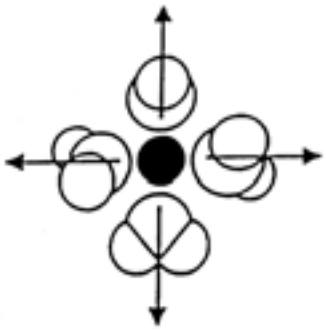


Figure 15. Sketch of preferential orientation of dipolar solvent molecules in the Coulombic field of small cations.

apparently irrotationally bound solvent molecules per cation or anion, respectively. For aqueous solutions the Z^+ values of some mono-, bi-, and trivalent cations [21] are collected in Fig. 17 where the diameter of dielectrically saturated water shells, relative to the ion diameter, is also shown graphically.

Among the monovalent ions only lithium and sodium induce a noticeable effect of dielectric saturation. With the alkaline earth metal ions saturation corresponds with roughly six completely irrotationally bound water molecules per cation and with trivalent main group aluminium, yttrium, and lanthanium ions twelve to thirteen water molecules per cation apparently do not contribute to the static permittivity of

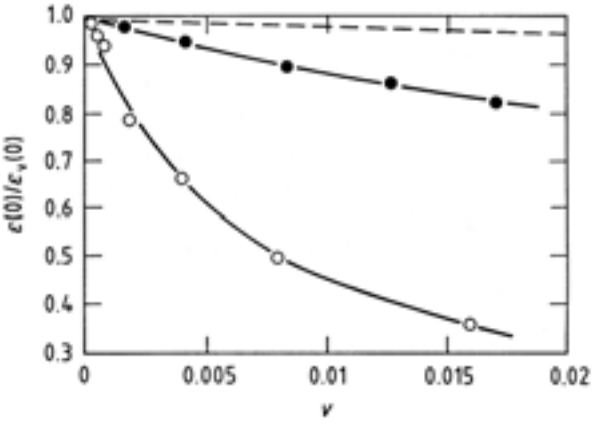


Figure 16. Permittivity ratio $\epsilon(0)/\epsilon_v(0)$ as a function of volume fraction v of salt for solutions of LiCl in water (• [77], $\epsilon_v(0) = \epsilon_w(0) = 78.35 \pm 0.05$ [46]) and in methanol (∘ [78], $\epsilon_v(0) = 32.64 \pm 0.20$ [46]) at 25°C. The dashed line represents the mixture relation (Eq. (16)).

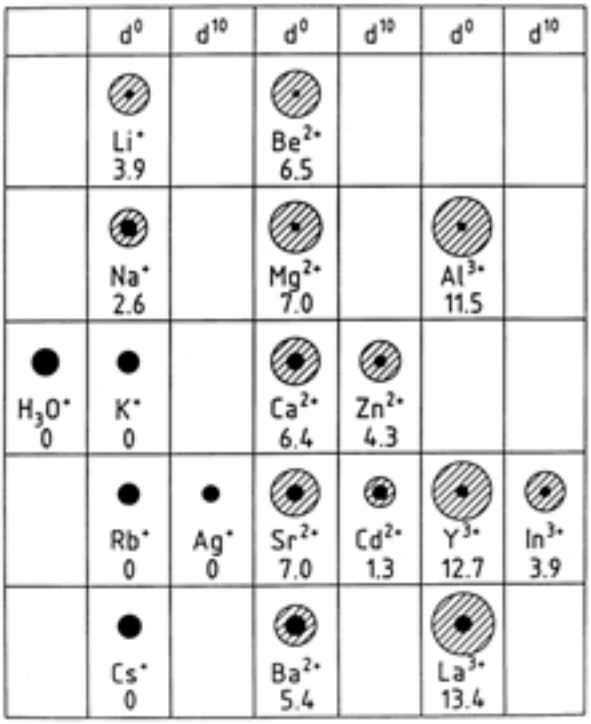


Figure 17. Numbers Z^+ of apparently irrotationally bound water molecules per ion for some mono-, di-, and trivalent cations [21] and graphical representation of the extent of dielectric saturation: hatched areas show the saturated water shell around the ions the crystallographic diameters of which are indicated by discs.

the solution. Quite remarkably the Z^+ values of the transition metal ions Zn^{2+}, Cd^{2+} and In^{3+} are considerably smaller than those of the similarly sized Ca^{2+}, Sr^{2+}, and Y^{3+} ions, respectively. These small Z^+ values have been taken an indication of cation-anion complex formation in the salt solutions containing transition metal ions. In vacuum the electronic configuration of such ions with complete d shell, like that of main group metal ions, involves a spherical charge distribution. In solution, however, interactions with anions are likely promoted by the directed d^{10} electron orbitals, thus increasing the tendency towards complex formation even with monovalent anions. The effect of dielectric saturation is smaller around the ion complex structures with reduced electric field strength than around separated ions with their comparatively strong Coulombic field. Therefore, the decrement in $\epsilon(0)$ is smaller in solutions with d^{10} cations than with equally sized d^0 cations. Complex formation in aqueous solutions of salts from transition metal cations and monovalent anions has been verified by solute contributions to the complex dielectric spectrum [21] and also by ultrasonic excess absorption spectra. An exceptional example is discussed in the next section.

4.2 Uncommon complexation of zinc(II)chloride

Similar to aluminium sulfate aqueous solutions (Fig. 12), the dielectric spectra of aqueous solutions of zinc chloride show two relaxation regions (Fig. 18). Again the high-frequency region reflects the reorientational motions of the solvent molecules. The additional low-frequency relaxation verifies in an obvious manner the existence of (dipolar) ion complex structures in transition metal halide aqueous solutions, as suggested from the ϵ_1 data of such systems (sect. 4.1). Because of the intriguing thermodynamic, structural, and transport properties of zinc chloride solutions these structures have been discussed for long time [86–88]. A multitude of different ion complexes has been identified using a variety of experimental techniques. In general the complex reaction scheme shown in Fig. 19 has to be considered when discussing

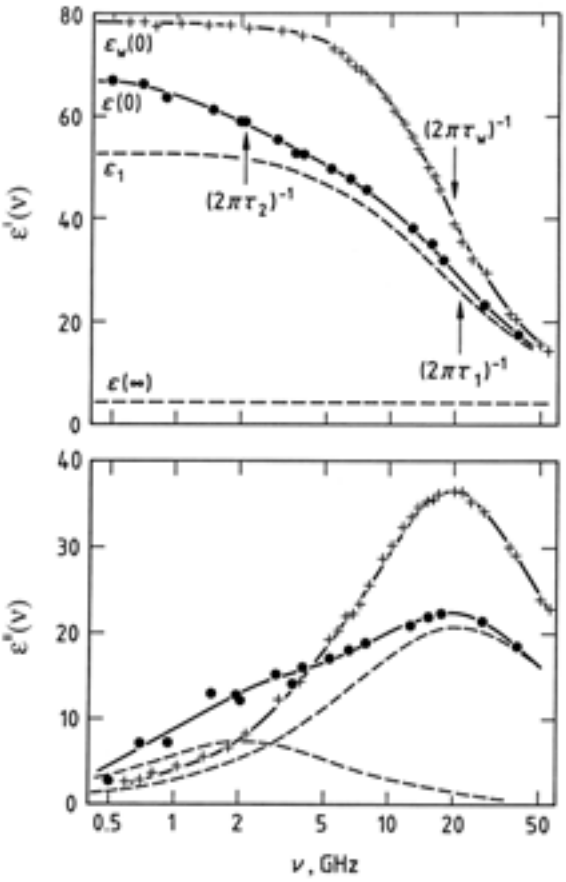


Figure 18. Real part $\epsilon'(\nu)$ and negative imaginary part $\epsilon''(\nu)$ of the dielectric spectrum of water (+) and of a 2.04 mol/l solution of $ZnCl_2$ in water (•) at 25°C [22,23]. Dashed lines show the subdivision of the spectrum for the salt solution according to Eq. (15) with $h_2 \equiv 0$.

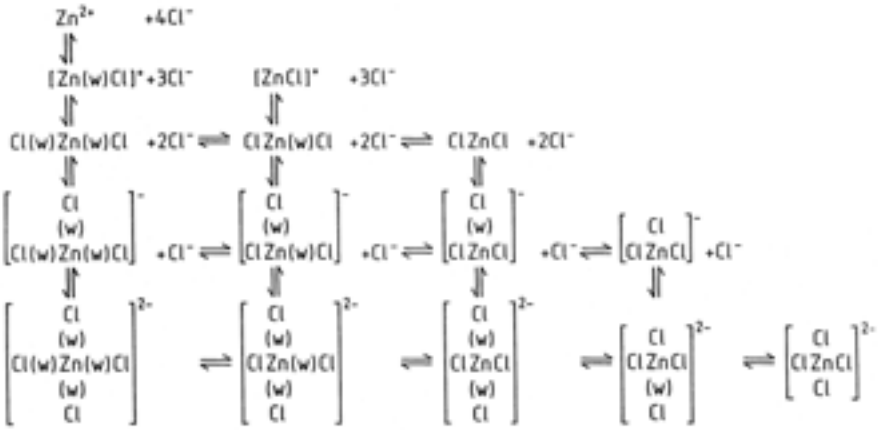


Figure 19. Zinc(II)chloride complexation scheme in water (w).

the characteristics of zinc chloride aqueous solutions. The formation of mono-, bi-, tri-, and tetrachloro complexes is well established now. All these species seem to exist not only as contact ion complexes but also as solvent separated outer-sphere species [89–92].

Following the above lines of reasoning (sect. 4.1) the extrapolated solvent contribution ϵ_1 to the static permittivity can be evaluated in terms of the concentration C_0 of completely dissociated zinc ions. According to our arguments zinc chloride ion complexes unlikely induce dielectric saturation effects. Let us assume the number Z^+ of apparently irrotationally bound water molecules around the completely dissociated Zn^{2+} ion to agree with that around the Mg^{2+} ion. The cation-water distances (Zn^{2+}: 2.08 Å; Mg^{2+}: 2.11 Å [93]) and the apparent molar volumes at infinite dilution (Zn^{2+}: -32.4 cm^3mol^{-1}; Mg^{2+}: -32.0 cm^3mol^{-1} [94]) almost agree with one another. The experimental ϵ_1 data can thus be evaluated in terms of the Zn^{2+} concentration. The relative Zn^{2+} content C_0/C following thereby is displayed in Fig. 20 where also predictions from a reasonable set of of equilibrium constants K_i, $i = 1, 2, 3, 4$ [95] for the association scheme of zinc ions and chloride ions are shown. In deriving this set of K_i values no distinction was made between inner sphere complexes and their water-containing analoga. The transition between these species and their outer-sphere analoga was assumed to be fast as compared to the transition between the complexes with different number of chloride ions. The agreement between our data from the ϵ_1 values and the predictions from the equilibrium constants of the multistep reaction scheme is striking. According to our expectations the relative content of completely dissociated Zn^{2+} ions decreases substantially with C.

Most species in the zinc chloride complexation scheme (Fig. 19) are nondipolar or only weakly dipolar. In the dichloro complex with linearly arranged chloride, zinc, and chloride ions, in the planar trichloro complex, as well as in the tetrahedrally structured tetrachloro complex the individual dipole moments largely compensate

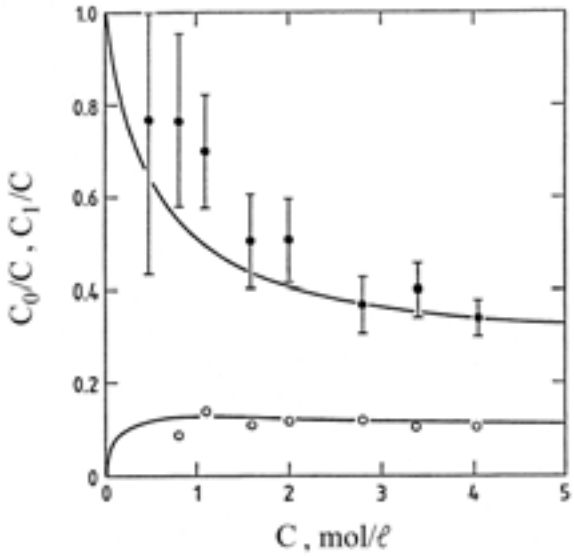


Figure 20. Concentration ratios $C_0/C(\bullet)$ and $C_1/C(\circ)$ of completely dissociated Zn^{2+} ions and of $(Zn(H_2O)Cl)^+$ complexes, respectively, as following from dielectric spectra of aqueous solutions of $ZnCl_2$ with salt concentration C [23]. The lines show the predictions from a set of equilibrium constants [95] for the zinc chloride association scheme.

each other. Therefore, the solute contribution $\epsilon(0) - \epsilon_1$ (Fig. 18) to the static permittivity of zinc chloride solutions appears to be mainly due to monochloro complexes. We used the relation

$$C_1 = (\epsilon(0) - \epsilon_1)\frac{\epsilon_0}{N}\left(\frac{2}{3}\right)^2 \frac{3k_BT}{\mu^2} \tag{18}$$

to derive the concentration C_1 of monochloro complexes from the relaxation amplitude $\epsilon(0) - \epsilon_1$ and the dipole moment μ of the ion pairs [23]. Here N is Avogadro's number and k_B is the Boltzmann constant. Using the simple relation

$$\mu = e \cdot l \tag{19}$$

for the dipole moment in solution, where e denotes the elementary charge and l the cation-anion distance, perfect agreement of the C_1 values with those from the set of equilibrium constants K_i, $i = 1, ..., 4$ is observed (Fig. 20), if $l = 4.2$ Åis assumed. This is the cation-anion distance in an outer-sphere complex. Hence, obviously, the monochloro complex exists predominantly as solvent-separated species.

The ultrasonic absorption spectra of zinc chloride solutions extend over a broader frequency band than a single Debye type relaxation term, thus indicating a distribution of relaxation times [30]. A relaxation time distribution had been already inferred from previous ultrasonic measurements, in a reduced frequency range, of zinc(II)halide aqueous solutions [96] and also from a combined ultrasonic and Bril-

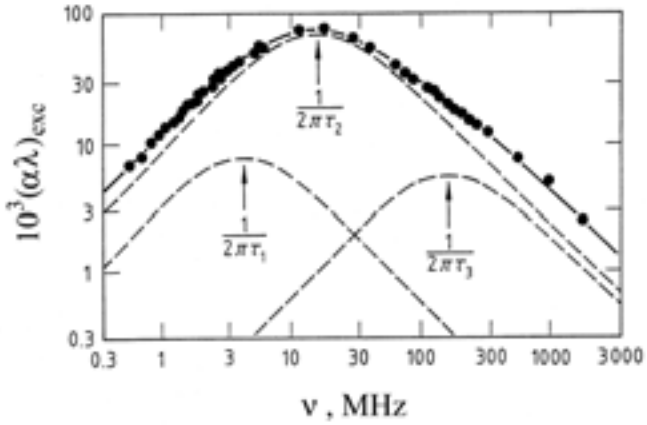


Figure 21. Ultrasonic excess absorption spectrum for a solution of 0.8 mol/l $ZnCl_2$ in water at 25°C. The subdivision of the spectrum into three Debye relaxation terms is indicated by dashed lines. The full line represents the sum of these terms [30].

louin scattering study of zinc chloride hydrated melts [97]. Three terms with discrete relaxation time are necessary to represent the experimental spectra adequately (Fig. 21). Consistency of the dependencies of relaxation parameters upon the salt concentration, however, requires even four Debye relaxation terms.

As a four-Debye-term model comprises too many unknown parameters the spectra have been evaluated assuming a reduced reaction scheme

$$\begin{aligned} Zn^{2+} + 4Cl^- &\rightleftharpoons (Zn^{2+})^* + 4Cl^- \rightleftharpoons (ZnCl)^+ + 3Cl^- \\ &\rightleftharpoons \quad ZnCl_2 + 2Cl^- \rightleftharpoons (ZnCl_3)^- + Cl^- \\ &\rightleftharpoons \quad (ZnCl_4)^{2-} \end{aligned} \tag{20}$$

in which $(Zn^{2+})^*$ is assumed an activated zinc ion, e.g. an ion with one hydration molecule less than Zn^{2+}. Relating the rate constants and reaction volumes of the coupled equilibrium to one another the number of unknown parameters in the regression analysis of the ultrasonic spectra has been reduced so that reliable rate constants and reaction volumes followed from the fitting procedures [30].

Alternatively, the multitude of proposed ion complex structures (Fig. 19) and the finding of precritical behaviour for concentrated zinc chloride aqueous solutions [98] has suggested the idea of rapidly fluctuating diffuse ion clusters rather than of stoichiometrically defined species [31]. The Romanov-Solov'ev model of noncritical concentration fluctuations [99–101] has been implemented. But, even if an additional Debye relaxation is taken into account, this model did not satisfactorily apply to the spectra at all concentrations. However, a superior analytical description of the experimental excess absorption spectra is reached with the Dissado-Hill ("DH") relaxation spectral function with only four unknown parameters (Fig. 22). This function has been originally derived for a uniform representation of a diversity of dielectric spec-

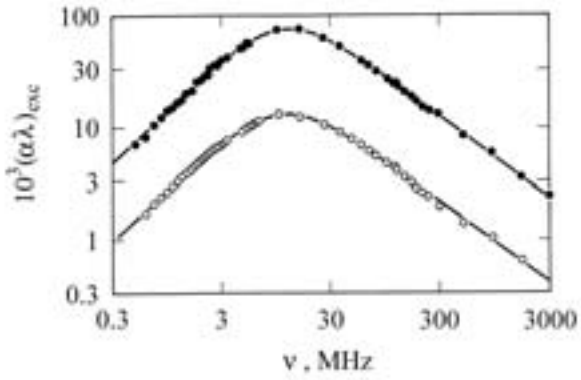


Figure 22. Ultrasonic excess absorption spectra for aqueous solutions of $ZnCl_2$ ($\circ$, 0.3mol/l; $\bullet$, 0.8mol/l) at 25°C. The lines are graphs of the Dissado-Hill function (Eq. (21)) with the parameter values given in Table 2.

tra [102,103]. Rewritten to apply to the sonic absorption per wavelength it reads

$$(\alpha\lambda)_{exc} \tag{21}$$
$$= A_{DH}\mathrm{Im}\left[\left(\frac{1}{1+i\omega\tau_{\mathrm{DH}}}\right)^{1-\mathrm{n_{DH}}} \lim_{\epsilon\to 0}\int_{\epsilon}^{1} \mathrm{t}^{-\mathrm{n_{DH}}}\left(1-\frac{\mathrm{t}}{1-i\omega\tau_{\mathrm{DH}}}\right)^{-(1-\mathrm{m_{DH}})}\mathrm{dt}\right].$$

The parameter values of the zinc chloride solutions are collected in Table 2. Within the framework of the Dissado-Hill model the ions are assumed to form clusters rather than well-defined complexes. Relaxations within the clusters are characterized by the principal relaxation time τ_{DH}. The parameter n_{DH} ($0 \leq n_{DH} \leq 1$) describes the correlations within the clusters. The limiting values $n_{DH} = 0$ and $n_{DH} = 1$ represent relaxation processes that are completely independent and strongly correlated,

C mol/l	A_{DH} 10^{-3}	τ_{DH} ns	n_{DH}	m_{DH}	B 10^{-12}s
0.1	1.4 ± 0.1	4.5 ± 0.3	0 ± 0.02	0.88 ± 0.02	32.6 ± 0.1
0.2	5.1 ± 0.5	14.0 ± 0.7	0.34 ± 0.03	0.90 ± 0.02	32.7 ± 0.2
0.3	18.9 ± 0.3	14.5 ± 0.3	0.25 ± 0.01	0.87 ± 0.01	33.7 ± 0.1
0.35	22.8 ± 0.7	18.0 ± 0.6	0.31 ± 0.02	0.90 ± 0.02	33.4 ± 0.2
0.4	36.8 ± 0.6	16.0 ± 0.4	0.24 ± 0.01	0.89 ± 0.01	33.6 ± 0.2
0.5	56.5 ± 1.1	15.3 ± 0.6	0.23 ± 0.02	0.85 ± 0.02	34.1 ± 0.2
0.6	85.7 ± 1.1	14.7 ± 0.3	0.20 ± 0.01	0.90 ± 0.01	34.6 ± 0.1
0.8	119.1 ± 2.5	14.3 ± 0.6	0.23 ± 0.02	0.94 ± 0.02	34.7 ± 0.4

Table 2. Parameters of the Dissado-Hill spectral function (Eq. (21)) for aqueous solutions of zinc(II) chloride at 25°C and salt concentration C [31].

respectively. The parameter m_{DH} ($0 \leq m_{DH} \leq 1$) reflects correlations between different clusters with $m_{DH} = 1$ constituting the limiting situation of ideally connected relaxation sequences in the cluster system. Such alternative considerations of the sound absorption spectra of electrolyte solutions have been currently renewed by a distribution function theory in which the sonic relaxations in the spectra are related to the long-range concentration fluctuations within the liquid [32]. We mention that the ultrasonic absorption spectra of 1:1 valent tetraalkylammonium bromide aqueous solutions exhibit strong indications of precritical fluctuations in the local concentration [33].

4.3 Other 2:1 valent salts

Present ultrasonic spectrometry is sufficiently sensitive to indicate the small-amplitude relaxations due to ion complex formation in aqueous solutions of 2:1 valent salts [34, 35,50]. For a transition-metal-chloride solution an example is given in Fig. 4. A spectrum with even smaller relaxation amplitude is shown in Fig. 23 where the ultrasonic excess absorption for a solution of an alkaline earth metal chloride is displayed versus frequency. All relevant excess absorption spectra measured so far can be well represented by a single Debye-type relaxation term (Eq. (7)). Parameters as resulting from a regression analysis of the spectra are given in Table 3.

The relaxation times τ of the solutions are strikingly similar. Obviously, the τ values do not noticeably depend upon the cation radius and on the electronic structure of the cation. As contact ion pair formation sensitively reflects the properties of the cations the observed relaxation is unlikely due to a process that involves in-

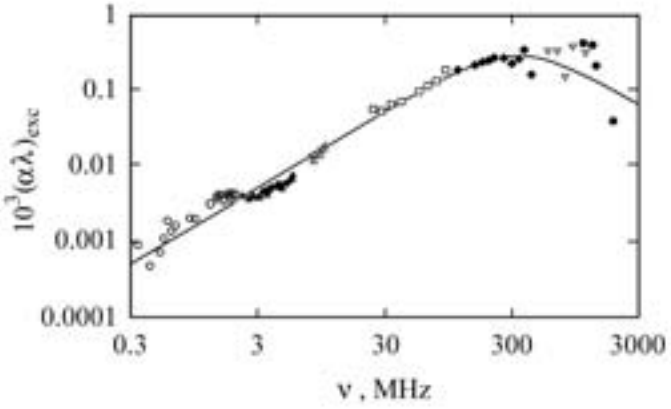


Figure 23. Ultrasonic excess absorption spectrum for a 1 mol/l solution of $SrCl_2$ in water at 25°C [35]. In order to disclose the excess absorption not to be due to systematic errors in the measurements, data from different specimen cells and apparatus are marked by figure symbols. The line represents a Debye relaxation term (Eq. (7)) with the parameter values given in Table 3.

Salt	Cec	r^+	C	c	A	A/C	τ	B
		10^{-10}m	mol/l	m/s	10^{-3}	cm^3mol^{-1}	ns	10^{-12}s
		[104]	±0.2%	±0.1%	±20%	±20%	±20%	±2%
				T=298.15 K				
$MgCl_2$	$2p^6$	0.66	1.0	1604.4	0.29	0.29	0.61	35.7
$CaCl_2$	$3p^6$	0.99	1.0	1587.0	0.59	0.59	0.37	35.8
$SrCl_2$	$4p^6$	1.12	1.0	1585.9	0.58	0.58	0.46	32.4
$NiCl_2$	$3p^6d^8$	0.69	0.8	1566.9	0.67	0.84	0.38	35.1
$CuCl_2$	$3p^6d^9$	0.72	0.5	1529.9	7.6	15.2	0.35	34.0
$Mg(NO_3)_2$	$2p^6$	0.66	1.0	1577.0	0.4	0.4	0.44	33.2
$Cu(NO_3)_2$	$3p^6$	0.99	1.0	1556.4	22.6	22.6	0.22	36.6
				T=283.15 K				
$Ca(NO_3)_2$	$3p^6$	0.99	0.2	1465.0	2.85	14.2	0.36	49.4
			0.5	1486.2	12.7	25.4	0.32	49.0
			1.0	1524.1	30.9	30.9	0.32	52.8
			1.5	1563.0	49.2	32.8	0.34	55.4
				C=1 mol/l				
			T, K					
$Ca(NO_3)_2$	$3p^6$	0.99	283.15	1524.1	30.9	30.9	0.32	52.8
			288.15	1536.2	28.8	28.8	0.28	45.5
			293.15	1547.0	25.0	25.0	0.25	40.6
			298.15	1556.4	22.6	22.6	0.22	36.6

Table 3. Parameters of the relaxation spectral function (Eqs. (6,7)) for ultrasonic absorption spectra of 2:1 valent salts [34,35]; Cec, cation electronic configuration; r^+, cation radius; c, sound velocity of solution.

ner sphere complexes. This argument is supported by previous results for the outer sphere-inner sphere complex equilibrium of sulfates in aqueous solutions, for which $\tau = 16\,\mu$s was found for Ni^{2+}, $\tau = 1.3\,\mu$s for Mg^{2+} and $\tau = 1.2$ ns only for Cu^{2+} (0.5 mol/l, 20°C [105]). For the chlorides $\tau = 0.38$ ns for Ni^{2+}, $\tau = 0.61$ ns for Mg^{2+}, and $\tau = 0.35$ ns for Cu^{2+} (0.5 mol/l≤C≤1 mol/l, 25°C, Table 3). Furthermore, the τ values for $Ca(NO_3)_2$ solutions at 10°C do not reveal a concentration dependence (Table 3), which is an indication of a unimolecular reaction. Estimation of the relaxation rate for the process of ion encounter, using the Debye-Eigen-Fuoss theory [106], predicts a relaxation term well above our frequency range of measurements. Since, on the other hand, the formation of outer-outer sphere complexes, which is even questioned with 2:2 valent electrolyte solutions (sect. 3.1), unlikely occurs in solutions of 2:1 valent salts, the Debye-type relaxation term has been assigned to the equilibrium

$$(M^{2+}\cdots L^{-})_{aq} \rightleftharpoons (M(H_2O)L)^{+}_{aq} \tag{22}$$

between the complex of encounter, $(M^{2+}\cdots L^{-})_{aq}$ and the outer-sphere complex.

The difference between the molar amplitudes A/C of the $CuCl_2$ solutions and the $NiCl_2$ solutions is striking. Like $NiCl_2$, solutions of the transition metal chlorides $MnCl_2$, $FeCl_2$, and $CoCl_2$ with $3p^6d^5$, $3p^6d^6$, and $3p^6d^7$, electronic structure, respectively, do not show noticeable ultrasonic excess absorption in the relevant frequency

range [50]. Ion complex formation of the subgroup metal ion Cu^{2+} seems to be stabilized by the Jahn-Teller effect. There are indications of an additional low-frequency relaxation term in the spectra of $CuCl_2$ solutions which has not been covered by the measurements because transducer electrodes of the resonator cells were corroded by the samples [50]. Likely this low-frequency term reflects the equilibrium between the ion pairs and the outer-sphere complexes.

Also striking is the substantial increase in the amplitude A of calcium salt solutions when chloride is substituted by nitrate (Table 3). We suppose stronger interactions of $(Ca^{2+})_{aq}$ with the nitrate ion, due to its lone pair electrons and its delocalized π orbital [104], than with the chloride ions with its spherical s^2p^6 electron configuration. An unexplained result is the considerably smaller difference in the amplitudes of $MgCl_2$ and $Mg(NO_3)_2$ solutions.

5 Conclusions

Multistep associations of ions have been among the first "immeasurable fast" reactions [1] that have been effectively measured and described in detail [9,10]. Nevertheless still today much interest is directed towards these reactions. Though the Eigen-Tamm mechanism has been verified by broadband ultrasonic spectrometry of scandium sulfate solutions [17] the existence of outer-outer sphere complexes in solutions of 2:2 valent salts is still under discussion [32]. As early evaluations of experimental spectra, without any assistance from computer facilities, used a slightly incorrect high-frequency background term in the ultrasonic excess absorption spectra, a distribution of relaxation times may have been simulated. More recent broadband spectra of $MnSO_4$ solutions (Fig. 9) reveal only two Debye-type relaxation terms within the frequency range of measurements. These terms are assigned to the coupled equilibria between completely dissociated ions and outer-sphere complexes and between the latter and contact ion pairs. Halides of second subgroup metals, such as zinc chloride, display a broad variety of ion complex structures in solution. Their spectra have indeed been discussed in terms of a complex equilibrium between stoichiometrically well-defined species [30] but have been also considered assuming fluctuating ion clusters with intra- and intercluster correlations [31]. Small-amplitude ultrasonic relaxation terms have been revealed for solutions of 2:1 valent electrolytes and have been related to the equilibrium between the complex of encounter and the outer-sphere complexes [34,35]. This equilibrium is of significance also for biochemistry as it may interfere with other association mechanisms, like counter ion condensation on polyelectrolytes, carbohydrate-cation interactions [107], as well as inclusion complex formation [108].

Acknowledgments

We are indebted to the technicians of the institute for their continual support with precision engineering components, electronic devices, computer facilities, and figure layout.

References

[1] M. Eigen, Nobel Lecture, 1967.

[2] K. Tamm and G. Kurtze, 'Absorption of Sound in Aqueous Solutions of Electrolytes', Nature **168**, 346 (1951).

[3] G. Kurtze, 'Untersuchung der Schallabsorption in wäßrigen Elektrolytlösungen im Frequenzbereich 3 bis 100 MHz', Nachr. Akad. Wiss. Gött. Math.-Phys.-Chem. Kl., 57 (1952).

[4] K. Tamm, 'Schallabsorptionsmessungen in Wasser und in wässerigen Elektrolytlösungen im Frequenzbereich 5 kHz bis 1 MHz', Nachr. Akad. Wiss. Gött. Math.-Phys.-Chem. Kl., 81 (1952).

[5] G. Kurtze and K. Tamm, 'Measurements of Sound Absorption in Water and in Aqueous Solutions of Electrolytes', Acustica **3**, 33 (1953).

[6] K. Tamm, G. Kurtze, and R. Kaiser, Measurements of Sound Absorption in Aqueous Solutions of Electrolytes, Acustica **4**, 280 (1954).

[7] K. Tamm, in *Handbuch der Physik 11.I*, edited by S. Flügge (Springer, Berlin, 1961).

[8] M. Eigen, G. Kurtze, and K. Tamm, 'Zum Reaktionsmechanismus der Ultraschallabsorption in wäßrigen Elektrolytlösungen', Z. Elektrochem. Ber. Bunsenges. Phys. Chem. **57**, 103 (1953).

[9] M. Eigen and K. Tamm, 'Schallabsorption in Elektrolytlösungen als Folge chemischer Relaxation I. Relaxationstheorie der mehrstufigen Dissoziation', Z. Elektrochem. Ber. Bunsenges. Phys. Chem. **66**, 93 (1962).

[10] M. Eigen and K. Tamm, 'Schallabsorption in Elektrolytlösungen als Folge chemischer Relaxation II. Meßergebnisse und Relaxationsmechanismen für 2-2-wertige Elektrolyte', Z. Elektrochem. Ber. Bunsenges. Phys. Chem. **66**, 107 (1962).

[11] K. G. Plaßand A. Kehl, 'Schallabsorption in Lösungen 2-2-wertiger Elektrolyte im Frequenzbereich 0.3 GHz bis 2.8 GHz', Acustica **20**, 360 (1968).

[12] L. G. Jackopin and E. Yeager, 'Ultrasonic Relaxation in Manganese Sulfate Solutions', J. Phys. Chem. **74**, 3766 (1970).

[13] K. Fritsch, C. J. Montrose, J. L. Hunter, and J. P. Dill, 'Relaxation Phenomena in Electrolytic Solutions', J. Chem. Phys. **52**, 2242 (1970).

[14] D. P. Fay and N. Purdi, 'Ultrasonic Absorption in Aqueous Salts of the Lanthanides III. Temperature Dependence of $LnSO_4$ Complexation', J. Phys. Chem. **74**, 1160 (1970).

[15] A. Bechtler, K. G. Breitschwerdt, and K. Tamm, 'Ultrasonic Relaxation Studies in Aqueous Solutions of 2-2-Electrolytes', J. Chem. Phys. **52**, 2975 (1970).

[16] P. Hemmer, 'The Volume Changes in Ionic Association Reactions', J. Phys. Chem. **76**, 895 (1972).

[17] A. Bonsen, W. Knoche, W. Berger, K. Giese, and S. Petrucci, 'Ultrasonic Relaxation Studies in Aqueous Solutions of Aluminium Sulphate and Scandium Sulphate', Ber. Bunsenges. Phys. Chem. **82**, 678 (1978).

[18] R. Pottel, 'Die komplexe Dielektrizitätskonstante wäßriger Lösungen einiger 2-2-wertiger Elektrolyte im Frequenzbereich 0.1 bis 38 GHz', Ber. Bunsenges. Phys. Chem. **69**, 363 (1965).

[19] R. Pottel, in *Chemical Physics of Ionic Solutions*, edited by B. E. Conway and R. G. Barradas (Wiley, New York, 1966).

[20] U. Kaatze and K. Giese, 'Dielectric Spectroscopy on Some Aqueous Solutions of 3:2 Valent Electrolytes. A Combined Frequency and Time Domain Study', J. Mol. Liq. **36**, 15 (1987).

[21] U. Kaatze, 'Dielectric Effects in Aqueous Solutions of 1:1, 2:1, and 3:1 Valent Elec-

trolytes: Kinetic Depolarization, Saturation, and Solvent Relaxation', Z. Phys. Chem. (Munich) **135**, 51 (1983).
[22] U. Kaatze, V. Lönnecke, and R. Pottel, 'Dielectric Spectroscopy on Aqueous Solutions of Zinc(II)Chloride. Evidence of Ion Complexes', J. Phys. Chem. **91**, 2206 (1987).
[23] U. Kaatze, V. Lönnecke, and R. Pottel, 'Dielectric Spectroscopy of Aqueous $ZnCl_2$ Solutions. Dependence upon Solute Concentration and Comparison with other Electrolytes', J. Mol. Liq. **34**, 241 (1987).
[24] R. Buchner, S. G. Capewell, G. Hefter, and P. M. May, 'Ion-Pair and Solvent Relaxation Processes in Aqueous Na_2SO_4 Solutions', J. Phys. Chem. B **103** 1185 (1999).
[25] R. Buchner, C. Hölzl, J. Stauber, and J. Barthel, 'Dielectric Spectroscopy of Ion-Pairing and Hydration in Aqueous Tetra-*n*-alkylammonium Halide Solutions', Phys. Chem. Chem. Phys. **4**, 2169 (2002).
[26] R. Buchner, F. Samani, P. M. May, P. Sturm, and G. Hefter,'Hydration and Ion Pairing in Aqueous Sodium Oxalate Solutions', ChemPhysChem **4**, 373 (2003).
[27] R. Buchner, T. Chen, and G. Hefter, 'Complexity in "Simple" Electrolyte Solutions: Ion Pairing in $MgSO_4$ (aq)', J. Phys. Chem. B **108**, 2365 (2004).
[28] T. Chen, G. Hefter, and R. Buchner, 'Ion Association and Hydration in Aqueous Solutions of Nickel(II) and Cobalt (II) Sulfate', J. Solution Chem. **34**, 1045 (2005).
[29] R. Buchner, W. W. Rudolph, and G. T. Hefter, 'Comment on "Dynamic Ion Association in Aqueous Solutions of Sulfate" ', J. Chem. Phys. **124**, 247101 (2006).
[30] U. Kaatze, and B. Wehrmann, 'Broadband Ultrasonic Spectroscopy on Aqueous Solutions of Zinc(II)Chloride. I. Kinetics of Complexation', Z. Phys. Chem. (Munich) **177**, 9 (1992).
[31] U. Kaatze, K. Menzel, and B. Wehrmann, 'Broadband Ultrasonic Spectroscopy on Aqueous Solutions of Zinc(II)Chloride. II. Fluctuations and Clusters', Z. Phys. Chem. (Munich) **177**, 27 (1992).
[32] T. Yamaguchi, T. Matsuoka, and S. Koda, 'Theoretical Study on the Sound Absorption of Electrolytic Solutions I: Theoretical Formulation', J. Chem. Phys., in press (2007).
[33] V. Kühnel, and U. Kaatze, 'Uncommon Ultrasonic Absorption Spectra of Tetraalkylammonium Bromides in Aqueous Solution', J. Phys. Chem. **100**, 19747 (1996).
[34] R. Behrends, P. Miecznik, and U. Kaatze, 'Ion-Complex Formation in Aqueous Solutions of Calcium Nitrate. Acoustical Spectrometry Study', J. Phys. Chem. A **106**, 6039 (2002).
[35] E. Baucke, R. Behrends, and U. Kaatze, 'Rapidly Fluctuating Ion Complexes in Aqueous Solutions of 2:1 Valent Salts', Chem. Phys. Lett. **384**, 224 (2004).
[36] F. Eggers and U. Kaatze, 'Broad-Band Ultrasonic Measurement Techniques for Liquids', Meas. Sci. Technol. **7**, 1 (1996).
[37] R. Behrends, K. Lautscham, and U. Kaatze, 'Acoustical Relaxation Spectrometers for Liquids', Ultrasonics **39**, 393 (2001).
[38] U. Kaatze and Y. Feldman, 'Broadband Dielectric Spectrometry of Liquids and Biosystems', Meas. Sci. Technol. **17**, R17 (2006).
[39] B. Voleišienė, G. Miglinienė and A. Voleišis, 'Ultrasonic Study of Ionic Association in Aqueous Solutions of Lanthanine Salts', J. Acoust. Soc. Am. **105**, 962 (1999).
[40] B. Voleišienė, G. Miglinienė, and A. Voleišis, 'Ultrasonic Study of the Complexation Kinetics of Aqueous Yttrium Nitrate Solutions', Chemija (Vilnius) **12**, 225 (2001).
[41] P. Debye, *Polar Molecules* (Chemical Catalog Co., New York, 1929).
[42] U. Kaatze, R. Behrends, and R. Pottel, 'Hydrogen Network Fluctuations and Dielectric Spectrometry of Liquids', J. Non-Crystalline Solids **305**, 19 (2002).
[43] U. Kaatze, in *Electromagnetic Aquametry, Electromagnetic Wave Interaction with*

Water and Moist Substances, edited by K. Kupfer (Springer, Berlin, 2005).
[44] U. Kaatze, 'Complex Permittivity of Water as a Function of Frequency and Temperature', J. Chem. Eng. Data **34**, 371 (1989).
[45] W. J. Ellison, K. Lamkaouchi, and J. M. Moreau, 'Water, a Dielectric Reference', J. Molec. Liquids **68**, 171 (1996).
[46] U. Kaatze, 'Reference Liquids for the Calibration of Dielectric Sensors and Measurement Instruments', Meas. Sci. Technol. **18**, 967 (2007).
[47] K. Nörtemann, J. Hilland, and U. Kaatze, 'Dielectric Properties of Aqueous NaCl Solutions at Microwave Frequencies', J. Phys. Chem. A **101**, 6864 (1997).
[48] U. Kaatze, 'Aspekte elektromagnetischer Aquametrie ionisch leitender Materialien', Technisches Messen **74**, 261 (2007).
[49] U. Kaatze and R. Behrends, 'Liquids: Formation of Complexes and Complex Dynamics', in *Oscillations, Waves, and Interactions. Festschrift on the occasion of 60 Years of Drittes Physikalisches Institut*, edited by T. Kurz, U. Kaatze and U. Parlitz (Universitätsverlag Göttingen, Göttingen, 2007) (this book).
[50] B. Wehrmann, *Ultraschallabsorptions und -dispersionsmessungen im Frequenzbereich von 500 kHz bis 3 GHz zur Frage der Komplexbildung in wäßrigen Lösungen von Zinkchlorid und Chloriden anderer 2-wertiger Metalle*, Dissertation, Georg-August-Universität Göttingen (1991).
[51] U. Kaatze, T. O. Hushcha, and F. Eggers, 'Ultrasonic Broadband Spectrometry of Liquids: A Research Tool in Pure and Applied Chemistry and Chemical Physics', J. Solution Chem. **29**, 299 (2000).
[52] U. Kaatze, R. Pottel, and A. Wallusch, 'A New Automated Waveguide System for the Precise Measurement of Complex Permittivity of Low-to-High-loss Liquids at Microwave Frequencies', Meas. Sci. Technol. **6**, 1201 (1995).
[53] U. Kaatze, V. Kühnel, K. Menzel, and S. Schwerdtfeger, 'Ultrasonic Spectroscopy of Liquids. Extending the Frequency Range of the Variable Sample Length Pulse Technique', Meas. Sci. Technol. **4**, 1257 (1993).
[54] U. Kaatze, V. Kühnel, and G. Weiss, 'Variable Pathlength Cells for Precise Hypersonic Spectrometry of Liquids up to 5 GHz', Ultrasonics **34**, 51 (1996).
[55] U. Kaatze and K. Lautscham, 'Below Cut-Off Piston Attenuator as Calculable Signal Vernier for Microwaves up to 15 GHz', J. Phys. E: Sci. Instrum. **19**, 1046 (1986).
[56] O. Göttmann, U. Kaatze, and P. Petong, 'Coaxial to Circular Waveguide Transition as High-Precision Easy-to-Handle Measuring Cell for the Broad Band Dielectric Spectrometry of Liquids', Meas. Sci. Technol. **7**, 525 (1996).
[57] A. P. Gregory and R. N. Clarke, 'Traceable Measurements of the Static Permittivity of Dielectric Reference Liquids over the Temperature Range 5-50°C', Meas. Sci. Technol. **16**, 1506 (2005).
[58] U. Kaatze, B. Wehrmann, and R. Pottel, 'Acoustical Absorption Spectroscopy of Liquids Between 0.15 and 3000 MHz: I. High Resolution Ultrasonic Resonator Method, J. Phys. E: Sci. Instrum. **20**, 1025 (1987).
[59] F. Eggers, U. Kaatze, K. H. Richmann, and T. Telgmann, 'New Plano-Concave Ultrasonic Resonator Cells for Absorption and Velocity Measurements in Liquids Below 1 MHz', Meas. Sci. Technol. **5**, 1131 (1994).
[60] R. Behrends, F. Eggers, U. Kaatze, and T. Telgmann, 'Ultrasonic Spectrometry of Liquids Below 1 MHz. Biconcave Resonator Cell with Adjustable Radius of Curvature', Ultrasonics **34**, 59 (1996).
[61] R. Polacek and U. Kaatze, 'A High-Q Easy-to-Handle Biconcave Resonator for Acoustic Spectrometry of Liquids', Meas. Sci. Technol. **14**, 1068 (2003).
[62] R. Hagen, R. Behrends, and U. Kaatze, 'Acoustical Properties of Aqueous Solutions

of Urea: Reference Data for the Ultrasonic Spectrometry of Liquids', J. Chem. Eng. Data **49**, 988 (2004).
[63] T. Yamaguchi, T. Matsuoka, and S. Koda, 'Theoretical Study on the Sound Absorption in Electrolytic Solutions II: Assignments of Relaxations', J. Chem. Phys., in press (2007).
[64] H. Strehlow, *Rapid Reactions in Solution* (VCH, Weinheim, 1992).
[65] F. Eggers and T. Funck, 'Ultrasonic Measurements with Millilitre Samples in the 0.5 – 100 MHz Range', Rev. Sci. Instrum. **44**, 969 (1973).
[66] K. S. Cole and R. H. Cole, 'Dispersion and Absorption in Dielectrics. I. Alternating Current Characteristics', J. Chem. Phys. **9**, 341 (1941).
[67] K. Tamm and M. Schneider, 'Bestimmung der komplexen DK von Elektrolytlösungen bei 0.3 – 2.5 GHz aus Dämpfungs- und Phasenmessungen mittels geregelter Dämpfungsleitung', Z. angew. Phys. **20**, 544 (1966).
[68] H. Falkenhagen, *Theorie der Elektrolyte* (Hirzel, Stuttgart, 1971).
[69] J. Barthel, H. Krienke, and W. Kunz, *Physical Chemistry of Electrolyte Solutions, Modern Aspects* (Steinkopff, Darmstadt, 1998).
[70] R. Buchner and J. Barthel, 'Dielectric Relaxation in Solutions', Annu. Rep. Progr. Chem. C **97**, 349 (2001).
[71] D. A. G. Bruggeman, 'Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. I. Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper aus isotropen Substanzen', Ann. Phys. (Leipzig) **5**, 636 (1935).
[72] K. Giese, U. Kaatze, and R. Pottel, 'Permittivity and Dielectric and Proton Magnetic Relaxation of Aqueous Solutions of the Alkali Halides', J. Phys. Chem. **74**, 3718 (1970).
[73] W.-Y. Wen and U. Kaatze, 'Aqueous Solutions of Azoniaspiroalkane Halides. 3. Dielectric Relaxation', J. Phys. Chem. **81**, 177 (1977).
[74] U. Kaatze and R. Pottel, 'Dielectric Properties of Organic Solute Water Mixtures. Hydrophobic Hydration and Relaxation', J. Mol. Liquids **52**, 181 (1992).
[75] U. Kaatze, 'Microwave Dielectric Properties of Liquids', Rad. Phys. Chem. **45**, 549 (1995).
[76] U. Kaatze, 'The Dielectric Properties of Water in Its Different States of Interaction', J. Solution Chem. **26**, 1049 (1997).
[77] R. Pottel, in: *Water, a Comprehensive Treatise, Vol.3: Aqueous Solutions of Simple Electrolytes*, edited by F. Franks (Plenum, New York, 1973).
[78] U. Kaatze, D. Adolph, D. Gottlob, and R. Pottel, 'Static Permittivity and Dielectric Relaxation of Solutions of Ions in Methanol', Ber. Bunsenges. Phys. Chem. **84**, 1198 (1980).
[79] U. Kaatze, M. Schäfer, and R. Pottel, 'The Complex Dielectric Spectrum of Aqueous Methanol and Isopropanol Solutions', Z. Phys. Chem. (Munich) **165**, 103 (1989).
[80] J. B. Hubbard, L. Onsager, W. M. van Beek, and M. Mandel, 'Kinetic Polarization Deficiency in Electrolyte Solutions', Proc. Natl. Acad. Sci. USA **74**, 401 (1977).
[81] J. B. Hubbard and L. Onsager, 'Dielectric Dispersion and Dielectric Friction in Electrolyte Solutions I.', J. Chem. Phys. **67**, 4850 (1977).
[82] J. B. Hubbard, 'Dielectric Dispersion and Dielectric Friction in Electrolyte Solutions II.', J. Chem. Phys. **68**, 1649 (1978).
[83] J. B. Hubbard, P. Colomonos, and P. G. Wolynes, 'Molecular Theory of Solvated Ion Dynamics III. The Kinetic Dielectric Decrement', J. Chem. Phys. **71**, 2652 (1979).
[84] P. G. Wolynes, 'Dynamics of Electrolyte Solutions', Ann. Rev. Phys. Chem. **31**, 345 (1980).
[85] B. J. Alder and W. E. Alley, 'Generalized Hydrodynamics', Phys. Today **37**, 56 (1984).

[86] D. E. Irish, in *Ionic Interactions. From Dilute Solutions to Fused Salts*, edited by S. Petrucci (Academic, New York, 1971).
[87] P. V. Giaquinta, M. P. Tosi, and N. H. March, 'Coordination Chemistry and Ionic Solvation in Divalent Metal Halide Aqueous Solutions', Phys. Chem. Liq. **13**, 1 (1983).
[88] H. Weingärtner, K. J. Müller, H. G. Hertz, A. V. J. Edge, and R. Mills, 'Unusual Behavior of Transport Coefficients in Aqueous Solutions of Zinc Chloride', J. Phys. Chem. **88**, 2173 (1984).
[89] D. E. Irish, B. McCaroll, and T. F. Young, 'Raman Study of Zinc Chloride Solutions', J. Chem. Phys. **39**, 3436 (1963).
[90] C. O. Quicksall and T. G. Spiro, 'Raman Spectra of Tetrahalozincates and the Structure of Aqueous $ZnCl_4^{2-}$', Inorg. Chem. **5**, 2232 (1966).
[91] G. S. Darbari, M. R. Richelson, and S. Petrucci, 'Ultrasonic Study of Aqueous Solutions of $ZnCl_2$: From Dilute Solutions to Hydrated Melts', J. Chem. Phys. **53**, 859 (1970).
[92] Y. Yongyai, S. Kokpol, and B. S. Rode, 'Microstructure and Species Distribution of Aqueous Zinc Chloride Solutions. Results from Monte Carlo Simulations', J. Chem. Soc. Faraday Trans. **88**, 1537 (1992).
[93] Y. Marcus, 'Ionic Radii in Aqueous Solutions', J. Solution Chem. **12**, 271 (1983).
[94] H. L. Friedman and C. V. Krishnan, in *Water, a Comprehensive Treatise, Vol.3: Aqueous Solutions of Simple Electrolytes*, edited by F. Franks (Plenum, New York, 1973).
[95] R. A. Horne, 'The Adsorption of Zinc(II) on Anion Exchange Resins. I. The Secondary Cation Effect', J. Phys. Chem. **61**, 1651 (1957).
[96] K. Tamura, 'Ultrasonic Absorption Studies of the Complex Formation of Zinc(II)Halides in Aqueous Solution', J. Phys. Chem. **81**, 820 (1977).
[97] R. Carpio, F. Borsay, C. Petrovic, and E. Yeager, 'Ultrasonic and Hypersonic Properties of Ionic Hydrate Melts', J. Chem. Phys. **65**, 29 (1976).
[98] V. S. Sperkach, private communication (1997).
[99] V. P. Romanov and V. A. Solov'ev, 'Sound Absorption in Solutions', Sov. Phys. Acoust. **11**, 68 (1965).
[100] V. P. Romanov and V. A. Solov'ev, 'Relaxation of the Ion Atmospheres and Sound Absorption in Electrolytes', Sov. Phys. Acoust. **19**, 550 (1974).
[101] V. P. Romanov and S. V. Ul'yanov, 'Bulk Viscosity in Relaxing Media', Phys. A **201**, 527 (1993).
[102] R. M. Hill, 'Characterisation of Dielectric Loss in Solids and Liquids', Nature **275**, 96 (1978).
[103] L. A. Dissado and R. M. Hill, 'The Fractal Nature of the Cluster Model Dielectric Response Functions', J. Appl. Phys. **66**, 2511 (1989).
[104] E. Arnold, *The Architecture and Properties of Matter. An Approach through Models* (Ormerod, M. B., London, 1970).
[105] S. Petrucci, in *Ionic Interactions. From Dilute Solutions to Fused Salts*, edited by S. Petrucci (Academic, New York, 1971).
[106] R. M. Fuoss, 'Ionic Association. III. The Equilibrium between Ion Pairs and Free Ions', J. Am. Chem. Soc. **80**, 5059 (1958).
[107] E. Baucke, R. Behrends, K. Fuchs, R. Hagen, and U. Kaatze, 'Kinetics of Ca^{2+} Complexation with Some Carbohydrates in Aqueous Solutions', J. Chem. Phys. **120**, 8118 (2004).
[108] J. Haller, P. Miecznik, and U. Kaatze, 'Ultrasonic Attenuation Spectrometry Study of α-Cyclodextrin + KI Complexation in Water', Chem. Phys. Lett. **429**, 97 (2006).

Oscillations, Waves and Interactions, pp. 367–404
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-14-1

Liquids: Formation of complexes and complex dynamics

Udo Kaatze[1] **and Ralph Behrends**[2]

[1]Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
[2]Fakultät für Physik, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. Acoustical relaxation spectra, measured in the frequency range between roughly 10^4 and $5 \cdot 10^9$ Hz, are discussed in view of the formation of mesoscopic molecular structures, like small complexes, stacks, and micelles, as well as of microheterogeneous liquid structures, as characteristic for noncritical and critical concentration fluctuations in binary systems. A variety of results is presented to show the capability of the method. An extended version of the model of micelle formation/decay kinetics is given that accounts for the special features of surfactant solutions near their critical micelle concentration. Also a unifying model of noncritical concentration fluctuations, that includes all previous theories, is shown to favourably apply the experimental findings. Evidence is presented indicating the need for a comprehensive theoretical treatment of systems revealing both, critical micelle formation and critical demixing properties.

1 Introduction

Liquids owe their fascinating and diverse features to molecular interaction energies on the order of the thermal energy. Thermal motions thus prevent liquids from establishing long-range order. Additionally, short-range order fluctuates rapidly. Let us consider water, the omnipresent chemical on our planet, as an example. Water is among the associating liquids. As the water molecule is capable of forming four hydrogen bonds, liquid water establishes a percolating three-dimensional hydrogen bond network. Due to thermal agitation a single hydrogen bond fluctuates with correlation time on the order of 0.1 to 1 ps [1]. Even the dielectric relaxation time of water, which reflects the period required for the reorientation of the dipolar molecules through a significant angle [2], is as small as 10 ps at room temperature [3]. Hence liquids are characterized by the rapid fluctuations of their short-range order. In order to understand liquid properties we need to investigate their molecular motions.

In addition to the establishment of the hydrogen bond network, mesoscopic molcular structures may be formed in aqueous solutions and in mixtures of water with other constituents. The knowledge of such structures and of their formation and decay processes is most important for our understanding of self-organization in liquids

with relevance to many phenomena in chemistry, physical chemistry and biochemistry, as well as for chemical engineering and process control. In aqueous systems a dominant factor in structure formation is the hydrophilic and hydrophobic interactions. The pressure exerted on hydrophobic parts of the non-aqueous constituents to reduce contact with water molecules leads to the formation of clusters, stacks, micelles, and bilayer membranes [4,5]. A variety of liquid mixtures minimizes energy by precritical or critical demixing [6–11]. Particularly exiting are systems forming molecular aggregates and simultaneously exhibiting demixing behaviour. Examples are solutions of amphiphiles which display a critical micelle concentration and also a critical demixing point [12–17].

In this review the molecular dynamics of such mesoscopic supramolecular liquid structures are discussed. We focus on evidence from broadband acoustical spectrometry as ourdays experimental techniques in that field are still based on pioneering work in Göttingen and especially also at the Dritte Physikalische Institut. We mention the benchmark papers by Kurtze, Tamm, and Eigen on the ultrasonic spectrometry of multivalent salt solutions revealing the stepwise dissociation of ions [18–20], by Plaß who was among the first to reach hypersonic measurement frequencies [21,22], and by Eggers who made the resonator method popular for liquid spectrometry [23]. Sonic waves which couple to thermodynamic parameters and to transport properties, such as the molecular volume and the shear viscosity, hold the potential to contribute to an elucidation of the formation of the complex structures mentioned above. Sonic signals probe the native systems. Therefore, no special labels or markers are necessary, as are required in many other sophisticated methods.

One of the advantages of acoustical spectrometry is the fact that thermal equilibrium of the sample under study is virtually kept during measurements, because only incremental perturbations result from pressure and temperature oscillations associated with the sonic waves. Another favourable feature is the almost universal character of the parameters interacting with sonic signals. The universal applicability of acoustical spetrometry, however, is connected to an often non-specific nature of results. It is therefore imperative to vary the systems to be investigated in a considered way, for example by variation of temperature, concentration, or solvent composition, or to supplement measurements with data from other methods.

2 Experimental aspects

2.1 Fundamentals of acoustical spectrometry

If the small loss from heat conduction is neglected the propagation constant

$$\gamma = \alpha + \imath\,\beta \tag{1}$$

of longitudinal waves in a liquid with density ϱ and shear viscosity η_s follows from the Navier-Stokes equation [24] as

$$\gamma^2 = \frac{-\omega\varrho}{K + \frac{4}{3}\imath\,\omega\eta_s}\,. \tag{2}$$

Here α is the attenuation coefficient, $\beta = 2\pi/\lambda$ is the wave number with $\lambda = c/\nu$ denoting the wavelength, c is the sound velocity, ν the frequency, and $\imath^2 = -1$. K is the complex adiabatic compression modulus, given by

$$K = \kappa_S^{-1}(0) + \imath\,\omega\eta_v\,, \tag{3}$$

where

$$\kappa_S(0) = \lim_{\omega\to 0}\left[-\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_S\right] \tag{4}$$

is the static adiabatic compressibility at very small angular frequency $\omega = 2\pi\nu$, and η_v is the volume viscosity. At very small attenuation ($\alpha \ll \beta$) the imaginary part of Eq. (2) yields

$$\alpha = \frac{2\pi^2\nu^2}{c^3\varrho}\left(\frac{4}{3}\eta_s + \eta_v\right). \tag{5}$$

Frequently it is assumed that, within the frequency range of measurement, roughly 10 kHz $\leq \nu \leq$ 10 GHz, the shear viscosity is independent of frequency but the volume viscosity

$$\eta_v(\nu) = \Delta\eta_v(\nu) + \eta_v(\infty) \tag{6}$$

may be composed of two parts, of which $\Delta\eta_v(\nu)$ displays relaxation characteristics whereas $\eta_v(\infty)$ does not depend upon frequency. The sonic attenuation coefficient

$$\alpha(\nu) = \alpha_{\mathrm{exc}}(\nu) + B'\nu^2 \tag{7}$$

thus contains a part with quadratic frequency dependence and coefficient

$$B' = \frac{2\pi^2}{c^3\varrho}\left(\frac{4}{3}\eta_s + \eta_v(\infty)\right) \tag{8}$$

and an excess contribution

$$\alpha_{\mathrm{exc}}(\nu) = \frac{2\pi^2}{c^3\varrho}\Delta\eta_v(\nu)\,, \tag{9}$$

which is of primary interest in acoustical spectrometry. Because of the frequency dependence of the asymptotic high-frequency "background" contribution (7) it is convenient to display experimental spectra in the frequency normalized format

$$\frac{\alpha}{\nu^2} = \frac{\alpha_{\mathrm{exc}}(\nu)}{\nu^2} + B'\,, \tag{10}$$

accentuating the low-frequency part of the spectrum. If interest is focussed on the high-frequency regime and if, particularly, comparison with theoretical models and their thermodynamic parameters is derived an $(\alpha\lambda)_{\mathrm{exc}}$-versus-$\nu$ plot is appropriate. It is obtained from subtracting the asymptotic high-frequency contribution

$$B\nu = B'c\nu \tag{11}$$

from the total attenuation per wavelength, $\alpha\lambda$.

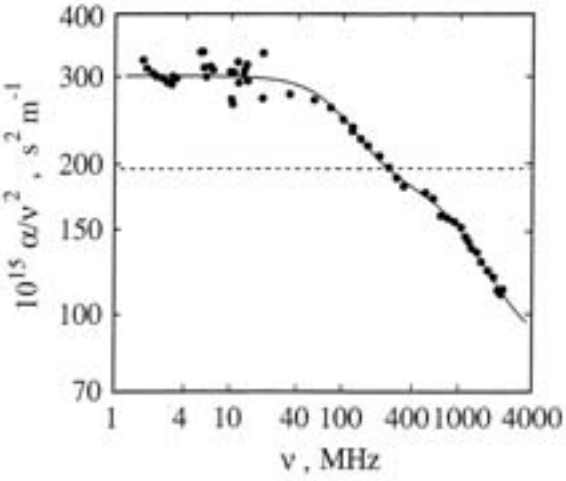


Figure 1. Ultrasonic attenuation spectrum in the frequency normalized format for *n*-dodecanol at 25°C [32]. The curve is the graph of a relaxation spectral function with two Debye-type relaxation terms. The dashed line shows the α/ν^2 value following from Eq. (12), assuming a frequency independent shear viscosity.

Constancy of the shear viscosity within the frequency range of measurements is not fulfilled with all liquids. Examples are polymer melts [26,27], the frequency dependent shear viscosity of which reflects modes of chain conformational isomerisation [27–30]. Another example is the chain isomerisation of *n*-alkanes [31] and alcohols [32]. Figure 1 shows a frequency normalized plot of the ultrasonic attenuation spectrum of *n*-dodecanol at 25°C. Also given by the dashed line is the contribution

$$\left(\frac{\alpha}{\nu^2}\right)_{\eta_{s0}} = \frac{8\pi^2}{3c^3\varrho}\eta_s(0) \tag{12}$$

that would result on assumption of a frequency independent shear viscosity $\eta_s(0)$ as measured with a capillary viscosimeter or a falling ball viscosimeter. The finding of experimental α/ν^2 data smaller than $\left(\alpha/\nu^2\right)_{\eta_{s0}}$ is a direct indication of the shear viscosity of *n*-dodecanol to be subject to a relaxation. Shear wave impedance spectrometry [33] has confirmed this conclusion by revealing a fequency dependent complex shear viscosity

$$\eta_s(\nu) = \eta_s'(\nu) - \imath\eta_s''(\nu)\,. \tag{13}$$

In Eq. (13) the real part $\eta_s'(\nu)$ represents the irreversible viscous molecular processes, whereas the negative imaginary part $\eta_s''(\nu)$ considers the reversible elastic mechanisms of the visco-elastic liquid. Figure 2 presents the shear viscosity of n-dodecanol, measured between some MHz and about 100 MHz, in a suggestive complex plane representation. The data evidently fit to the semicircular arc which is given as the graphical representation of the Debye-type relaxation spectral function [34]

$$R_s(\nu) = \eta_s(\infty) + \frac{A_s}{1 + \imath\,\omega\tau_s} \tag{14}$$

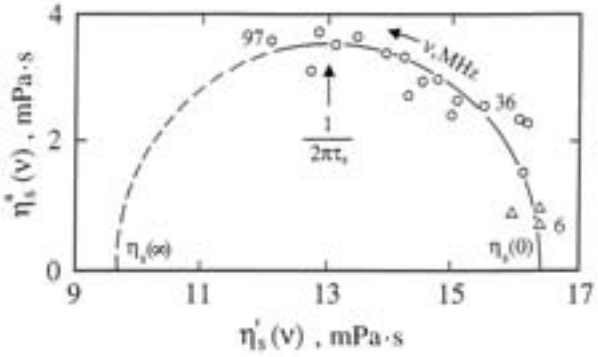


Figure 2. Complex plane representation of the frequency dependent shear viscosity of n-dodecanol at 25°C [32]. Figure symbols indicate data from different shear impedance resonator cells. The circular arc is a plot of the spectral function defined by Eq. (14) with the values for the parameters $\eta_s(0)$, $\eta_s(\infty)$, and τ_s found by a nonlinear regression analysis.

with discrete relaxation time. In this function $\eta_s(\infty)$ is the extrapolated high-frequency shear viscosity and $A_s = \eta_s(0) - \eta_s(\infty)$ is the relaxation amplitude. Hence, in general, a frequency dependent shear viscosity has to be taken into account by using

$$\alpha(\nu) = \frac{2\pi\nu^2}{c^3\varrho}\left(\frac{4}{3}\eta_s(\nu) + \eta_v(\nu)\right) \tag{15}$$

as the more general version of Eq. (5).

2.2 Attenuation spectrometry

Within the frequency range from 12 kHz to 4.6 GHz, which is currently available for the acoustical spectrometry of liquids, the wavelength λ of the sonic field within the sample varies by a factor of about $4 \cdot 10^5$. Due to the quadratic frequency dependence of the asymptotic high-frequency term in α (Eq. (7)) the variation in the attenuation coefficient is even as large as $(4 \cdot 10^5)^2 = 1.6 \cdot 10^{11}$. It is thus impossible to cover the frequency range with only one method of measurement. The spectra discussed in this review have been obtained applying two different techniques and numerous specimen cells, each one matched to a frequency range and to the sample properties, in order to reach a maximum sensitivity and to reduce experimental errors to as small as possible values.

At low frequencies ($\nu \lesssim 20$ MHz), where normally α is small, the resonator principle is appropriate as it is based on convoluting the acoustical path via multiple reflections, hereby increasing the effective pathlength of interaction with the sample. Calibration measurements using a reference liquid with carefully adjusted sound velocity and density are necessary in order to correct the measured data for intrinsic resonator loss. Spherical resonator cells [35] as well as cylindrically shaped cavities for quasi-one-dimensional wave propagation are in use. Popular are biplanar cavities with both

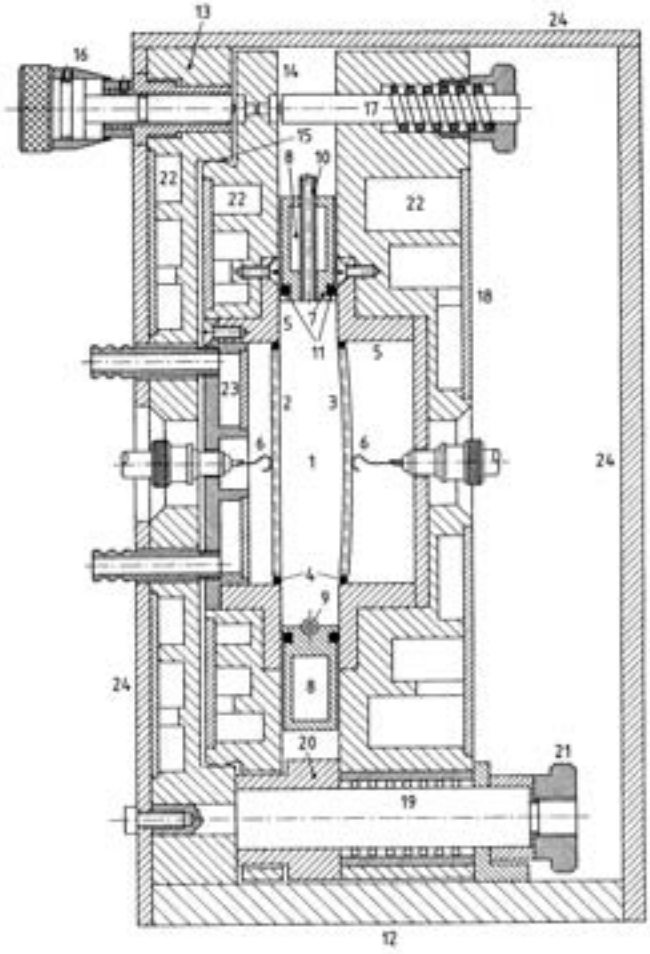


Figure 3. Cross section of a plano-concave resonator cell [37]. 1, sample volume; 2, planar circular piezoelectric transducer crystal with coaxially evaporated films of chrome and gold constituting the electrodes; 3, transducer disc as 2 but concavely shaped with radius of curvature =2 m; 4, layer of silicone rubber with embedded bronze strips providing electrical contact between the transducer front and the holding frame; 5, crystal setting; 6, spring contact; 7, cell jacket with 8, channel for circulating thermostat fluid; 9, inlet and 10, outlet for the sample liquid; 11, sealing O-ring; 12, base plate; 13, main frame fixed with respect to 12; 14, adjustable frame with 15, ball joint; 16, precisely adjustable screw with 17, counteracting spring; 18, movable frame fixed against radial dispacements and tilting by 19, ball-bush guides; 20, gauge block establishing the distance between 13 and 18; 21, locking device; 22 and 23, thermostatic channels; 24, thermostatic shield.

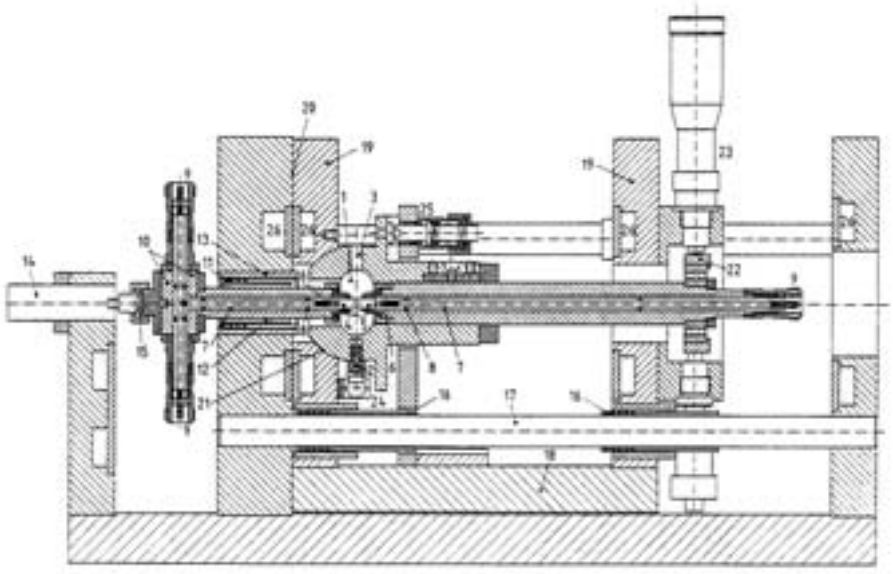


Figure 4. Construction of a variable-path-length cell for measurements between 0.6 and 4.6 GHz [42]. 1, cavity for sample cell with thermostatic jacket (not shown); 2 and 3, hole for inlet and outflow of sample; 4, sapphire rod serving as delay line and holding the piezoelectric zinc oxide transducer layer on its rear face; 5, mount for 4; 6, spring clamped electrical contact; 7, coaxial line; 8, plastic disc supporting the inner conductor of 7; 9, N-type connector; 10, T-branch; 11, high precision ball-bush guide; 12, specially honed and lapped bush; 13, bush; 14, piezo-translator; 15, backlash-free joint; 16, as 11; 17, finely honed and lapped pin; 18, sliding carriage; 19, mounting plate; 20, finely polished reference plane; 21, ball-and-socket joint; 22, adjustment device; 23, differential screws for sensitive adjustment of the sapphire rod direction; 24, ball gudgeon preventing the sample cell from rotating; 25, spring for the ball-and-socket joint 21; 26, thermostatic channels.

faces formed by piezoelectrical transducers [36]. In order to reduce diffraction losses plano-concave [37] and biconcave devices [38,39] are also employed. As an illustration, a detail drawing of a plano-concave cell is shown in Fig. 3. In many applications the focussing effect of concavely shaped faces includes further favourable features, such as a reduction of undesired effects from disturbances due to small changes in the resonator adjustment on variation of temperature and superior mode spectrum as compared to the biplanar cell.

In the upper frequency range ($\nu \gtrsim 10\,\mathrm{MHz}$) absolute α measurements are enabled by transmitting pulse-modulated sonic waves through a cell of variable sample length. The specimen cells mainly differ from one another by piezoelectric transmitter and reciever unit and by their sample volume [40–42] which, because of the smaller wavelength, may be much smaller at high frequencies. As an example the construction of the variable-path-length cell for hypersonic measurements between 0.6 and 4.6 GHz is presented in Fig. 4.

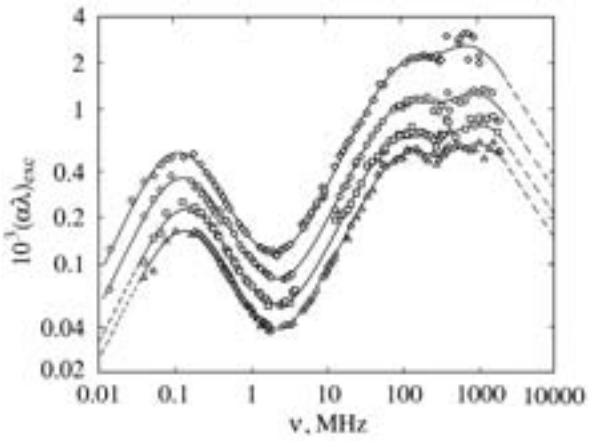


Figure 5. Sonic excess attenuation spectra for aqueous solutions of D-fructose at 25°C [44]. The saccharide concentrations are △, 0.5 mol/l; □, 0.7 mol/l; ∘, 1 mol/l; ⋄, 1.5 mol/l. The lines represent a relaxation spectral function with three Debye-type relaxation terms and with parameter values from a nonlinear fitting procedure.

Basically ourdays principles of measurement are the same as in early acoustical relaxation studies [43]. More sophisticated liquid spectrometry is enabled by a higher precision of the cell constructions, by advanced electronics, and by computer controlled measuring routines. Mechanical stability offers a 10 nm resolution in the cell length, which is mandatory for measurements in the GHz region where the wavelength is on the order of 300 nm and where the extraordinary large attenuation coefficient allows for small variations in the sample length only [42]. Modern electronics enable progressive measurement routines, scanning the complete transfer function of resonators to properly consider effects from higher order modes and running routinely calibrations in the variable-path-length method. Automatic measuring procedures finally facilitate multiple measurements for the reduction of statistical errors. Examples of broadband attenuation spectra are shown in Fig. 5.

2.3 Sound velocity dispersion

According to the Kramers-Kronig relations acoustical attenuation originating from relaxation processes is associated with a dispersion in the sound velocity c, which could be also utilized for liquid spectrometry. Normally, however, the dispersion in c is small and it is notoriously difficult to reach, in a significant frequency range, a sufficiently high accuracy in the sound velocity measurements. But since the sound velocity follows as a byproduct from the attenuation coefficient spectrometry, it can be used to consider the dispersion corresponding with large sonic absorption. Figure 6 shows the dispersion in c that is related to the formation of ion complexes in an aqueous solution of zinc chloride [45]. The total dispersion step of that system amounts to $c(\infty) - c(0) = 0.01 \cdot c(0)$ which should not be neglected in the evaluation procedures.

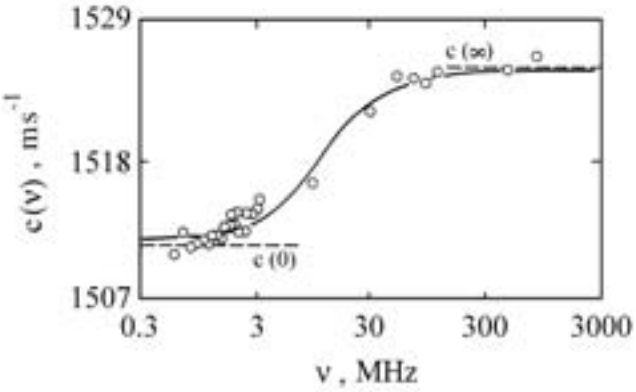


Figure 6. Sound velocity spectrum of a solution of 0.4 mol/l $ZnCl_2$ in water at 25°C [45]. The full line represents the dispersion in c as calculated from the corresponding excess attenuation spectrum.

3 Complexes and aggregates

3.1 Small complexes, stacks

Addition of calcium salts to carbohydrate solutions leave spectra of some systems virtually unchanged, whereas significant effects result for others, among them solutions of D-fructose and D-xylose [46] as well as methyl-β-D-arabinopyranoside and 1,6-anhydro-β-glucopyranoside [47]. Figure 7 indicates the substantial effect of $CaCl_2$ on the sonic excess attenuation spectrum of a solution of D-xylose in water. These changes in the ultrasonic spectra are assigned to Ca^{2+} complex formations with the carbohydrate (ch). Interactions between carbohydrates and cations are ubiquituous in nature and are believed to be significant in biochemistry. A detailed analysis of the acoustical spectra of carbohydrate solutions with added calcium salts reveals a stepwise complexation mechanism [46–48], in accordance with the Eigen-Winkler model [49]

$$(\mathrm{ch})_{\mathrm{aq}} + \left(\mathrm{Ca}^{2+}\right)_{\mathrm{aq}} \rightleftharpoons \left(\mathrm{ch} \cdots \mathrm{Ca}^{2+}\right)_{\mathrm{aq}} \rightleftharpoons \left(\mathrm{ch} - \mathrm{Ca}^{2+}\right)_{\mathrm{aq}} \rightleftharpoons \left(\mathrm{ch} \equiv \mathrm{Ca}^{2+}\right)_{\mathrm{aq}}. \quad (16)$$

In this model $\left(\mathrm{ch} \cdots \mathrm{Ca}^{2+}\right)_{\mathrm{aq}}$ denotes a solvent-separated outer-sphere complex, $\left(\mathrm{ch} - \mathrm{Ca}^{2+}\right)_{\mathrm{aq}}$ is a monodentate contact complex in which the cation interacts with the lone electrons of a carbohydrate ring, and $\left(\mathrm{ch} \equiv \mathrm{Ca}^{2+}\right)_{\mathrm{aq}}$ is a tridentate complex in which the cation interacts with lone electrons of three carbohydrate oxygens as sketched in Fig. 8.

Another example of carbohydrate complexation is the formation of cyclodextrin inclusion complexes. Cyclodextrins are cyclic glucosyl oligomers in which the monosaccharide rings are α-(1,4) linked. The cycloamyloses consist of six, seven or eight glucopyranose units (α-, β-, or γ-cyclodextrin, respectively), forming a truncated cone with a more hydrophobic inner cavity [50]. The latter allows the cyclic oligomers to

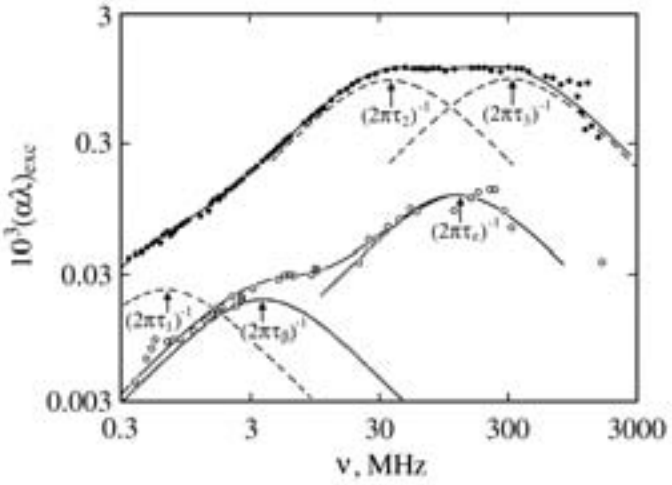


Figure 7. Ultrasonic excess attenuation spectra for aqueous solutions of 1 mol/l D-xylose at 25°C without $CaCl_2$ (○) and with 1 mol/l $CaCl_2$ added (●). Dotted and dashed lines indicate the subdivision of the former and the latter spectrum, respectively, in Debye-type relaxation terms. Solid lines represent the sum of these terms, respectively [46].

form inclusion complexes with a variety of molecules [51–55] and ions [56–60], including drugs [61,62]. Due to the hydrophilic outer surface of the cyclodextrin cone the complexes with guest molecules are soluble in water. Cyclodextrins are thus used to enhance or provide the solubility of molecules or organic ions, they act as stabilizers, selective agents, molecular recognition systems and also as capsules for the controlled delivery of specific molecules, with wide fields of applications.

Recently α-cyclodextrin (α-CD)-iodide (I^-) complex formation has been found [60] that follows the simple equilibrium

$$\alpha - \mathrm{CD} + \mathrm{I}^- \underset{k^r}{\overset{k^f}{\rightleftharpoons}} \alpha - \mathrm{CD} \cdot \mathrm{I}^- \tag{17}$$

where $\alpha - \mathrm{CD} \cdot \mathrm{I}^-$ denotes the monoiodide inclusion complex and k^f and k^r are the forward and reverse rate constants, respectively, related to one another by the

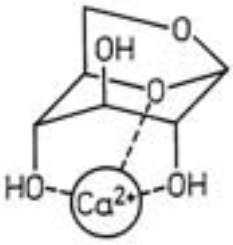


Figure 8. Structure of a tridentate calcium ion-anhydroglucopyranoside complex.

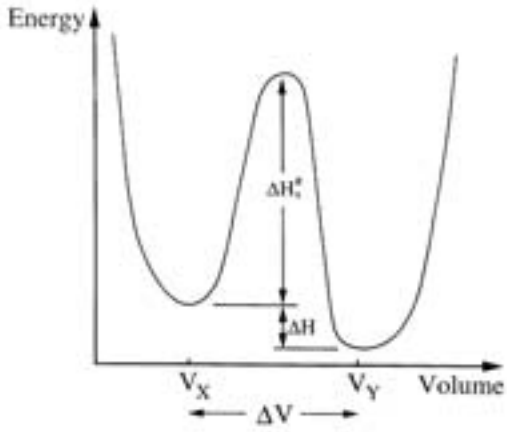


Figure 9. Scheme of relaxation process associated with a chemical equilibrium between species X and species Y.

equilibrium constant

$$K = k^f/k^r. \tag{18}$$

Within the framework of the simple reaction scheme shown in Fig. 9, the species on both sides of Eq. (17) differ from one another by the volume difference ΔV and the enthalpy difference ΔH. They are separated from one another by the enthalpy barriers $\Delta H_X^\ddagger$ and $\Delta H_Y^\ddagger = \Delta H_X^\ddagger + \Delta H$, where normally $\Delta H \ll \Delta H_X^\ddagger$, thus $\Delta H_Y^\ddagger \approx \Delta H_X^\ddagger$. Here X and Y denote $\alpha - \mathrm{CD} + \mathrm{I}^-$ and $\alpha - \mathrm{CD} \cdot \mathrm{I}^-$, respectively. The van't Hoff equation

$$\frac{d \ln K}{d T^{-1}} = -\frac{\Delta H}{R} \tag{19}$$

relates the equilibrium constant to the enthalpy difference ΔH, indicating that, according to our expectations, the larger the reaction enthalpy ΔH the larger K, that is the more the equilibrium is shifted to the right-hand side of Eq. (17).

The thermal activation relaxation scheme of Fig. 9 predicts ultrasonic excess attenuation that features Debye-type relaxation characteristics

$$(\alpha\lambda)_{\mathrm{exc}} = R_D(\nu) = \frac{A\omega\tau}{1+\omega^2\tau^2} \tag{20}$$

with the relaxation time τ given by

$$\tau^{-1} = k^f \left([\alpha - \mathrm{CD}] + \left[\mathrm{I}^-\right] + K^{-1} \right) \tag{21}$$

and the relaxation amplitude following as

$$A = \frac{\pi \Gamma c(\infty)}{RT} \Delta V_S^2 \,. \tag{22}$$

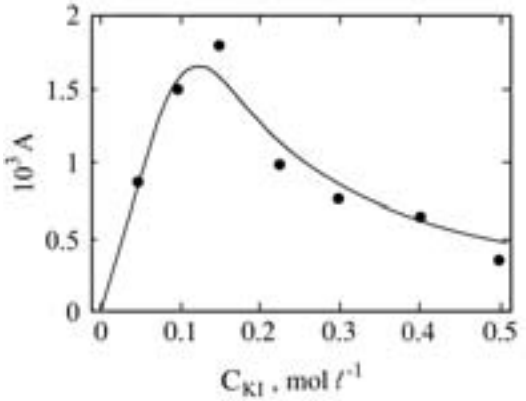


Figure 10. Amplitude of the sonic relaxation term reflecting the complex formation in aqueous solutions of 0.1 mol/l α-cyclodextrin with potassium iodide at 25°C displayed versus salt concentration C_{KI} [60].

Here Γ is a stoichiometric factor, which for the above equilibrium is given by

$$\Gamma^{-1} = [\alpha - \mathrm{CD}]^{-1} + \left[\mathrm{I}^-\right]^{-1} + \left[\alpha - \mathrm{CD} \cdot \mathrm{I}^-\right]^{-1} , \tag{23}$$

and ΔV_S denotes the adiabatic reaction volume

$$\Delta V_S = \frac{\Lambda_\infty}{\varrho\, C_{p\infty}} \Delta H - \Delta V \tag{24}$$

with the limiting high-frequency values Λ_∞ and $C_{p\infty}$ of the thermal expansion coefficient and the specific heat at constant pressure, respectively. Taking into account that the concentrations of the uncomplexed and complexed species are related to one another by the total cyclodextrin and iodide concentrations, the amplitude A of the ultrasonic relaxation term reflecting scheme (17) can be well described by relations (22) and (23) if the reasonable values $K = 50\,(\mathrm{mol/l})^{-1}$ and $\Delta V_S = 5.4\,\mathrm{cm}^3/\mathrm{mol}$ are assumed (Fig. 10).

Another interesting association mechanism is the formation of stacks from purine bases, assumed to be relevant in biology, because it constitutes the dominant interaction maintaining the secondary structure of biopolymers, such as DNA. Purine bases are polyaromatic ring molecules with hydrophilic sites mainly at the periphery and largely hydrophobic upper and lower faces. In order to prevent these faces from contact with water, the disc shaped bases form stacks in aqueous solutions. As with amphiphilic surfactants condensing to micelles, the self-aggregation is controlled by an isodesmic (sequential) reaction scheme

$$\mathrm{N}_1 - \mathrm{N}_i \underset{k_i^r}{\overset{k_i^f}{\rightleftharpoons}} \mathrm{N}_{i+1} \ , \ i = 1, 2, \ldots \tag{25}$$

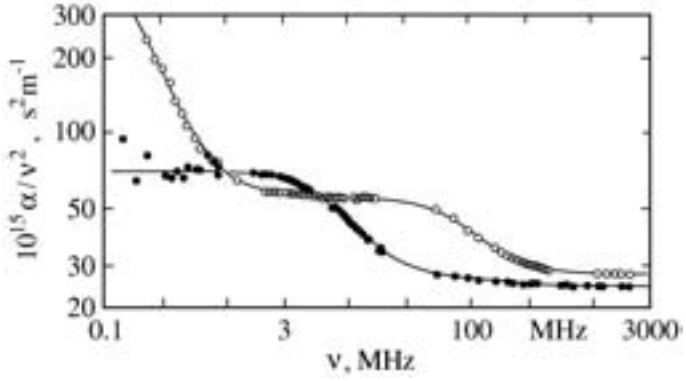


Figure 11. Frequency normalized ultrasonic spectra for aqueous solutions of 6-methylpurine without HCl (∘, 0.6 mol/l, pH=6.8) and with HCl added (•, 0.55 mol/l, pH=2) at 20°C [64].

where N_i denotes a multimer made of i monomers. For the stack formation, however, the shape of the size distribution function of the aggregates is different from that of micelles, particularly as there are no geometrical restrictions for the stack size [63]. Figure 11 shows ultrasonic attenuation spectra for aqueous solutions of 6-methyl-purine without and with HCl added [64]. The significant effect of pH upon the attenuation data below 3 MHz suggests the low-frequency relaxation term to be due to the proton exchange of the ampholytic 6-methylpurine molecules. The term at higher frequencies has been assigned to the stacking of the polyaromatic molecules. Unfortunately no clear evidence resulted for a preference of the sequential isodesmic reaction scheme (Eq. (25)) or the random isodesmic scheme

$$N_i + N_j \underset{k_{ij}^r}{\overset{k_{ij}^f}{\rightleftharpoons}} N_{i+j} \quad , i,j = 1,2,\dots , \tag{26}$$

because, within the concentration range available in the measurements, both models predict similar concentration dependencies for the relaxation times.

3.2 Micelles

The micelle formation/decay kinetics of nonionic surfactant solutions forming almost globularly shaped micelles with mean aggregation number $\overline{m}$ larger than about 50 (*proper* micelles), typically corresponding with critical micelle concentration cmc smaller than 10^{-2} mol/l, can be also well described by the sequential isodesmic scheme of coupled reactions defined by Eq. (25). It is assumed that the monomer concentration $[N_1]$ is much higher than that of any aggregate so that direct association of oligomers according to Eq. (26) with $i, j > 1$ can be neglected. Aniansson and Wall [65,66] have introduced a symmetrically bell-shaped nearly Gaussian size distribution

$$\left(k_{i+1}^r \left[\overline{N}_{i+1}\right] - k_i^r \left[\overline{N}_i\right]\right) / \left[\overline{N}_i\right] = -\left(i - \overline{m}\right) k^r/\sigma^2 \tag{27}$$

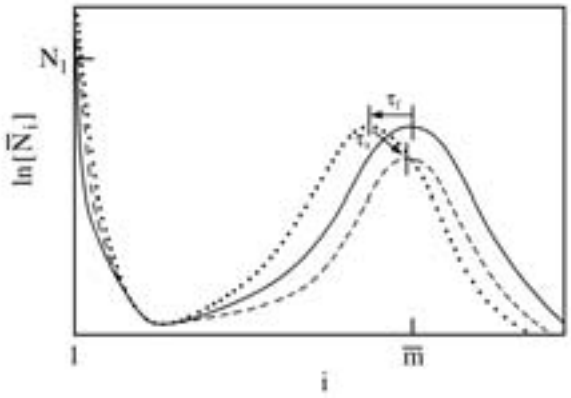


Figure 12. Distribution of the equilibrium concentration $\left[\bar{\mathrm{N}}_i\right]$ of aggregates from i monomers for proper micelle systems (full line). Dashed and dotted lines show the slow and fast response, respectively, of the surfactant system to external disturbances.

as sketched in Fig. 12. Here $\left[\overline{\mathrm{N}}_i\right]$ denotes the equilibirium concentration of species N_i, k^r is the mean reverse rate constant for micelle sizes around the mean $\bar{m}$ and σ^2 is the variance of the size distribution. Using Eq. (27) it is assumed that each step in the series of reactions is characterized by a well defined free energy change

$$\Delta G_i = RT \ln (K_i) \tag{28}$$

as following from the van't Hoff equation (Eq. (19)). Here $\Delta G_i = \Delta H_i - T\Delta S_i$ and $K_i = k_i^f / k_i^r$.

On the basis of the Aniansson-Wall model, the Teubner-Kahlweit theory [67,68] predicts sonic spectra with two Debye-type relaxation terms. These terms can be identified in a suggestive manner with two modes in the reformation of the equilibrium distribution of aggregates after a small disturbance. As indicated in Fig. 12, the fast relaxation process, with relaxation time τ_f roughly in the range of nanoseconds and microseconds, is due to the monomer exchange. It is characterized by a change of the aggregation number $\overline{m}$ at almost constant number of micelles per volume. In parallel, a slow process with relaxation time τ_s on the order of milliseconds or seconds proceeds by which the final equilibirium between the micelles and the suspending phase is reached. This slow process has been studied by time domain methods, predominantly by pressure jump and temperature jump techniques [69]. Here we discuss the low-frequency relaxation term of experimental spectra (Fig. 13 [70]) in the light of the fast monomer exchange. The term at even higher frequencies is assigned to the rotational isomerisation of the alkyl chains within the micellar cores.

For nonionic proper micelle systems with large $\overline{m}$, the monomer concentration $\left[\overline{\mathrm{N}}_1\right]$ is usually identified with the cmc. Using the scaled concentration

$$x = (C - \mathrm{cmc}) / \mathrm{cmc} \tag{29}$$

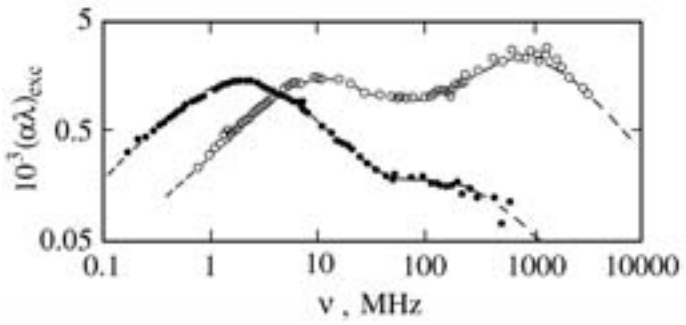


Figure 13. Ultrasonic excess attenuation spectra of aqueous solutions of *n*-decyltrimethylammonium bromide at 25°C at surfactant concentration 0.15 mol/l (•) and 0.5 mol/l (∘). The critical micelle concentration is 0.06 mol/l [70].

with the total amphiphile concentration $C = \sum_i i \left[\overline{\mathrm{N}}_i\right]$, the amplitude and relaxation time of the term reflecting the fast monomer exchange are predicted as [67,68]

$$A_f = \frac{\pi \left(\Delta V\right)^2 \mathrm{cmc}}{\kappa_{S\infty} R\, T} \frac{\sigma^2}{\overline{m}} x \left(1 + \frac{\sigma^2}{\overline{m}} x\right)^{-1} \tag{30}$$

and

$$\tau_f^{-1} = \frac{k^r}{\sigma^2} \left(1 + \frac{\sigma^2}{\overline{m}} x\right) , \tag{31}$$

respectively. In deriving Eq. (30) the same reaction volume ΔV has been assumed for all reaction steps of the isodesmic scheme. The quantity $\kappa_{S\infty} = \varrho^{-1} c^{-2}(\infty)$ is the adiabatic compressibility extrapolated to frequencies well above the relaxation region. The amplitude of the Teubner-Kahlweit-Aniansson-Wall model increases monotonously with concentration to asymptotically approach

$$\lim_{C \to \infty} A_f = \frac{\pi \left(\Delta V\right)^2 \mathrm{cmc}}{\kappa_{S\infty} R\, T} . \tag{32}$$

The relaxation rate depends linearly upon concentration

$$\tau_f^{-1} = aC + b \tag{33}$$

with slope $a = k^r / (\overline{m}\,\mathrm{cmc})$ and with $b = k^r \left(\sigma^2 - \overline{m}^{-1}\right)$. As obvious from Fig. 14, A_f and τ_f^{-1} for sodium dodecylsulfate solutions display quite different concentration dependencies [71], indicating a significant effect from the ionic nature of the amphiphile. Obviously, the discrepancy between the experimental findings and the predictions of the original Teubner-Kahlweit-Aniansson-Wall model reflects considerable effects from incomplete dissociation of the surfactant [72]. If J_i denotes the number of undissociated monomers per aggregate of class i, the law of mass action relates the concentration of free counterions

$$[\mathrm{N}_c] = \left(\frac{[\mathrm{N}_{i+1}]}{[\mathrm{N}_i]\,[\mathrm{N}_1]\,b_{i+1}}\right)^{1/(J_{i+1} - J_i)} \tag{34}$$

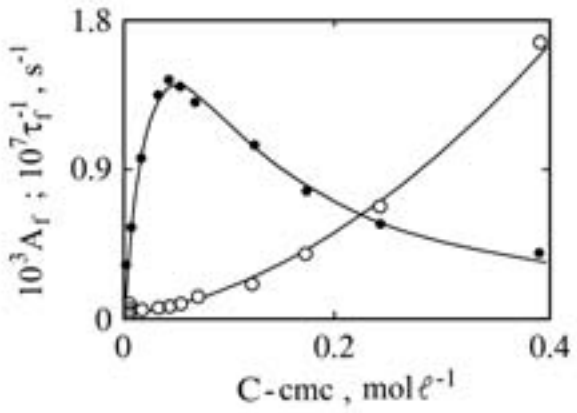


Figure 14. Relaxation amplitude A_f (•) and relaxation rate τ_f^{-1} (○) of the sonic relaxation term reflecting the fast monomer exchange in sodium dodecylsulfate solutions in water at 25°C [71] displayed versus concentration exceeding the cmc. Lines indicate the theoretical relations (Eqs. (30,31)) when according to Eq. (35) the monomer concentration $[N_1]$ is used instead of the cmc in the scaled concentration x (Eq. (29)).

to the concentration $[N_1]$ of surfactant monomers and $[N_i]$ of class i aggregates. Here b_{i+1} is the equilibrium constant. Assuming the effective degree of dissociation $\alpha_i = 1 - J_i/i$ for proper micelles to be independent of i within the micelle region ($\alpha_i = \alpha_{\overline{m}}$), an implicit relation for the monomer concentration as a function of total surfactant concentration C follows:

$$[N_1] = N_\gamma \left(\frac{C}{[N_1]} - 1 \right)^{1/(\overline{m}(2-\alpha_{\overline{m}})} \left[1 + \alpha_{\overline{m}} \left(\frac{C}{[N_1]} - 1 \right) \right]^{(1-\alpha_{\overline{m}})/(2-\alpha_{\overline{m}})} \tag{35}$$

where

$$N_\gamma = \left(\sqrt{2\pi} \sigma \overline{m} b_{\overline{m}} \right)^{1/[\overline{m}(2-\alpha_{\overline{m}})]} . \tag{36}$$

Deriving these equations $\overline{m} \approx \overline{m} + 1$ has been tacitely assumed. Using the monomer concentration $[N_1]$ from Eq. (35) instead of the cmc in the scaled concentration parameter x (Eq. (29)) and assuming the reasonable degree of dissociation $\alpha_{\overline{m}} = 0.33$ [73] the Teubner-Kahlweit-Aniansson-Wall theory nicely represents the experimental sonic amplitudes and relaxation times. The monomer concentrations following from the analysis of spectra are shown in Fig. 15 as a function of surfactant concentration. Also presented in that diagram is the cmc and the $[N_1]$-versus-C dependence for nonionic surfactant systems, as simulated by $\alpha_{\overline{m}} = 1$ in the above relations. With the ionic surfactant solutions the monomer concentration after reaching the cmc decreases significantly whereas the $[N_1]$ values of the nonionic surfactant system slightly increase with C. The tendency in the monomer concentration of ionic surfactants to decrease above the cmc is a well-established fact [5].

Another extension of the Teubner-Kahlweit model is required to properly account for the ultrasonic attenuation spectra of surfactant systems close to the cmc [74–

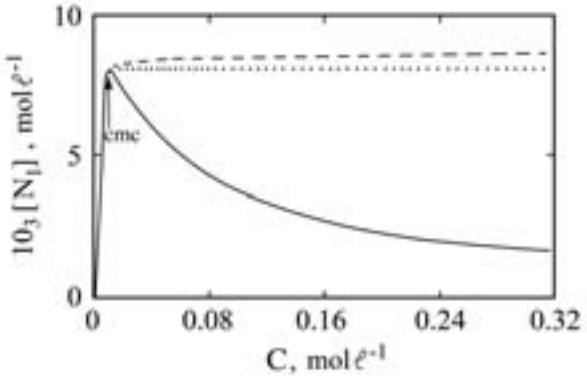


Figure 15. Monomer concentration $[N_1]$ of sodium dodecylsulfate solutions as resulting from the evaluation of the ultrasonic relaxation amplitudes (Fig. 14) shown as a function of surfactant concentration C. The dashed line shows $[N_1]$ on the assumption of completely dissociated surfactant molecules ($\alpha_{\overline{m}} = 1$), the dotted line indicates $[N_1] = \mathrm{cmc}$ [71].

76]. As an example, spectra for aqueous solutions of *n*-heptylammonium chloride (*n*-HepACl) at surfactant concentrations slightly above and slightly below the cmc (≈0.45 mol/l) are displayed in Fig. 16. Amphiphiles with such short alkyl chain do not reveal a well defined sharp critical micelle concentration. Rather they display a transition region. Obviously, even at surfactant concentrations somewhat below the transition region, both relaxation ranges of the proper sodium dodecylsulfate micelle system (Fig. 13) exist also in the spectra of the short chain *n*-HepACl solutions with extraordinarily high cmc (Fig. 16). Some features in the spectra of the latter,

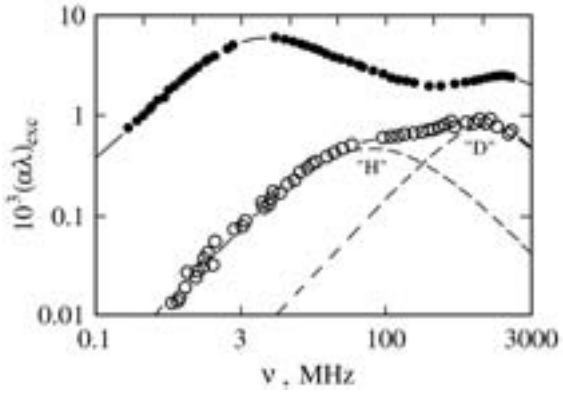


Figure 16. Sonic excess attenuation spectra for aqueous solutions of *n*-heptylammonium chloride at 25°C and at two concentrations: ○, 0.4 mol/l ≈ cmc; ●, 0.5 mol/l [74]. The dashed lines show the subdivision of the former spectrum in a Hill term (H) and a Debye term (D). Full lines are the graphs of the sum of these terms, respectively.

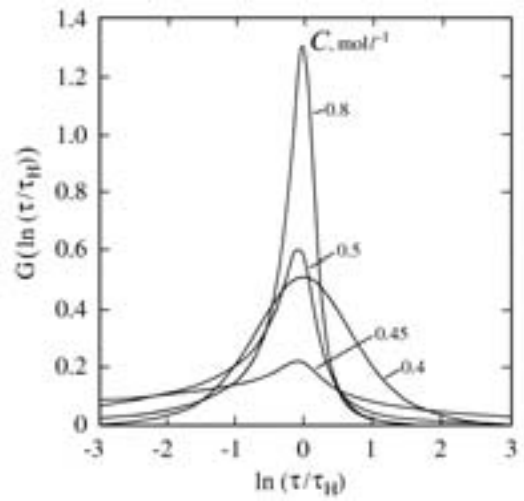


Figure 17. Relaxation time distribution function of the Hill relaxation term of some *n*-heptylammonium chloride aqueous solutions at 25°C.

however, attract attention. Obviously, the low-frequency term, which within the micelle formation/decay scheme (Fig. 12) reflects the fast monomer exchange, is subject to a considerable distribution of relaxation times. Empirically this term can be adequately represented by a Hill relaxation function [77–79], defined by

$$R_H(\nu) = \frac{A_H(\omega\tau_H)^{m_H}}{[1+(\omega\tau_H)^{2s_H}]^{\frac{m_H+n_H}{2s_H}}}\,. \tag{37}$$

In this function A_H is an amplitude and m_H, n_H, $s_H \in]0,1]$ are parameters controlling the shape and width of the underlying relaxation time distribution function $G_H(\ln(\tau/\tau_H))$. The principal relaxation time τ_H, according to [79]

$$\tau_H = \tau_{\max}(m_H/n_H)^{1/(2s_H)} \tag{38}$$

is related to the frequency $\nu_{\max} = (2\pi\tau_{\max})^{-1}$ at which $R_H(\nu)$ adopts its maximum.

At some surfactant concentrations the relaxation time distribution function, defined by

$$R_H(\nu) = A_H \int_{-\infty}^{\infty} G_H\left(\ln(\tau/\tau_H)\right)\omega\tau\left(1+(\omega\tau)^2\right)^{-1} d\ln(\tau/\tau_H) \tag{39}$$

is shown in Fig. 17 for solutions of *n*-HepACl in water. The function $G_H\left(\ln(\tau/\tau_H)\right)$ has been calculated by analytical continuation [80] from the Hill spectral function (Eq. (37)) using the normalisation

$$\int_{-\infty}^{\infty} G_H\left(\ln(\tau/\tau_H)\right) d\ln(\tau/\tau_H) = 0\,. \tag{40}$$

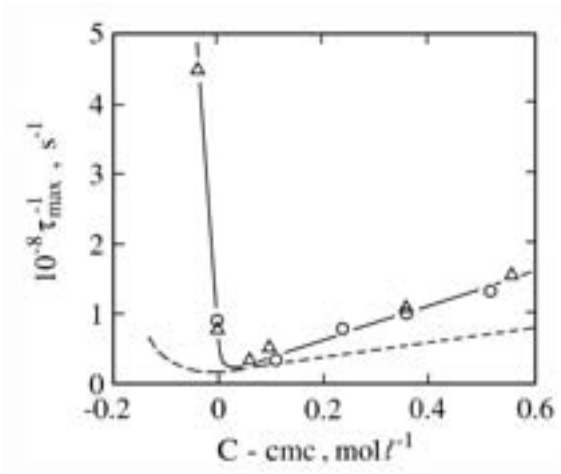


Figure 18. Relaxation rate $\tau_{\max}^{-1}$ of the Hill relaxation term in the spectra of aqueous solutions of n-heptylammonium chloride ($\triangle$ [74]) and of triethylene glycol monohexyl ether ($\circ$ [81]) at 25°C displayed versus concentration difference $C - \mathrm{cmc}$. The dashed line represents the predictions from the extended version of the Teubner-Kahlweit-Aniansson-Wall model [75].

The curves given in Fig. 17 reveal a particularly broad distribution function at surfactant concentrations near the cmc, where the content of oligomeric structures is high. Additionally, the relaxation time τ_H of the short chain surfactant system shows a remarkable behaviour (Fig. 18). At variance with the predictions of the Teubner-Kahlweit-Aniansson-Wall theory (Eq. (31)), the experimental relaxation rate $\tau_{\max}^{-1}$ (and thus τ_H^{-1}) at surfactant concentrations near the cmc first decreases with C to increase according to Eq. (31) at higher C only. In order to take properties of short chain surfactant systems into account, a computer simulation study of the coupled isodesmic reaction scheme (Eq. (25)) has been performed [75] in which the size distribution of micelles was not introduced empirically but was derived from reasonable rate constants, assumed to follow

$$k_{i+1}^f = k_{\overline{m}}^f \left(1 - s_f(i - \overline{m})\right) \tag{41}$$

and

$$k_{i+1}^r = \left[k_{\overline{m}}^r \left(1 + s_r(i - \overline{m})\right) + k_2^r \left(1 + \exp\left(\frac{1 - i_c}{d}\right)\right)\right] \Big/ \left(1 + \exp\left(\frac{i - i_c}{d}\right)\right) . \tag{42}$$

In these equations the parameters s_f and s_r define the slopes in the dependencies of k_i^f and k_i^r upon i and $k_{\overline{m}}^f$ as well as $k_{\overline{m}}^r$ allow the forward and reverse rate constants, respectively, to be matched at $i = \overline{m}$. Parameter i_c defines the aggregation number at which the reverse rate constants k_i^r change from a linear dependence upon i to a Fermi distribution. The quantities k_2^r and d are additionally used to model the k_i^r-versus-i relation at small aggregation numbers. These quantities are thus related to the cmc

of the system and, on favourable conditions, can be obtained from experimental data. At reasonable values of the parameters of Eqs. (41,42) a Gaussian size distribution of micelles follows also from this procedure. According to

$$\sigma^2 = (s_f + s_r)^{-1} \tag{43}$$

the variance of the distribution is given by the slope parameters in the dependencies of the rate constants upon the aggregate size.

The adequacy of the above relations has been demonstrated by considering results for solutions of proper micelles [75]. As the numerical evaluation of the isodesmic reaction scheme (Eq. (25)) avoids approximations of the analytical treatment, it is capable of revealing some special features of short chain surfactant solutions which remain unnoticed otherwise. For parameters modelling the n-HepACl system the size distribution of the micellar species at some surfactant concentrations is displayed in Fig. 19. Some characteristics of these distribution functions attract attention. The less pronounced relative minimum in the size distribution $\left[\overline{\mathrm{N}}_i(i)\right]$ contrasts the deep minimum presumed by the Teubner-Kahlweit-Aniansson-Wall model for the oligomer region. Hence there exists a noticeable content of small oligomeric structures in the short chain surfactant solutions and the separation between the slow and the fast relaxation process is reduced or absent at all. At $C \approx$ cmc and $C <$ cmc the relative maximum in $\left[\overline{\mathrm{N}}_i(i)\right]$, as characteristic for the Gaussian distribution of proper micelles, is absent. Even at surfactant concentrations distinctly smaller than the cmc, a noticeable content of oligomeric species is formed.

The eigenrate spectrum of the isodesmic reaction scheme (Eq. (25)) near the cmc confirms the absence of any slow relaxation process due to the absence of a pronounced relative minimum in the size distribution $\left[\overline{\mathrm{N}}_i(i)\right]$. As indicated by the

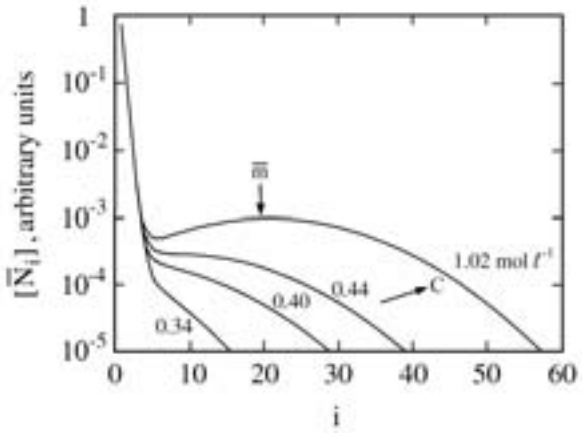


Figure 19. Distribution of the equilibrium concentration $\left[\overline{\mathrm{N}}_i\right]$ of aggregates from i monomers as resulting from the extended Teubner-Kahlweit model [75]. In the numerical calculations parameters have been chosen to correspond with the n-heptylammonium chloride/water system at surfactant concentration C.

dashed line in Fig. 18 the extended model of stepwise association predicts an increase of the relaxation rate $\tau_{\max}^{-1}$ when, at C below the cmc, the surfactant concentration decreases. An increase distinctly smaller than the one obtained from sonic attenuation spectra, however, is predicted by the extended Teubner-Kahlweit model. The theoretical relaxation-rate-versus-concentration relation also suffers from the inappropriate assumption of a concentration independent mean aggregation number $\overline{m}$, leading to a wrong slope $d(\tau_{\max}^{-1})/dC$ at $C >$ cmc. Interesting, however the numerical evaluation of the isodesmic reaction scheme yields the simultaneous presence of two fast relaxation terms with similar relaxation rates and relaxation amplitudes. These terms cannot be represented by a single Debye term but result in an unsymmetric broadening of the relaxation time spectrum as indicated by the Hill relaxation time distribution. Furtheron the numerical simulation reveals ultrafast relaxation contributions reflecting the monomer exchange of oligomeric species. This oligomer process is assigned to the high-frequency Debye-type relaxation term in the sonic attenuation spectra of short chain surfactant solutions because this term is unlikely due to the hydrocarbon chain isomerisation in the micelle cores. Structural isomerisations of such short chains are expected to display relaxation characteristics to the sonic attenuation spectrum at frequencies well above measurement range.

4 Local fluctuations in concentration

4.1 Noncritical dynamics

Binary liquid mixtures may minimize their free energy by forming a microheterogeneous structure which fluctuates rapidly in time. The local fluctuations in the concentration of the constituents relax by diffusion. Based on the dynamic scaling hypothesis [82–85]

$$\tau_\xi = \frac{\xi^2}{2D}\,, \tag{44}$$

the characteristic relaxation time τ_ξ is assumed to be given by a characteristic length ξ of the system and by the mutual diffusion coefficient D. In critical mixtures the fluctuation correlation length ξ covers vast ranges of size and follows a power law [6–11]

$$\xi = \xi_0 \epsilon^{-\tilde{\nu}} \tag{45}$$

thus tending to mask the individual properties of the system. In Eq. (45) ξ_0 is an individual amplitude, $\tilde{\nu}$ a universal exponent and

$$\epsilon = \frac{|T - T_c|}{T_c} \tag{46}$$

is a scaled (reduced) temperature. In systems displaying noncritical dynamics the critical temperature is not reached so that the fluctuation correlation length does not diverge.

Ultrasonic attenuation spectra due to noncritical concentration fluctuations [86–90] extend over a broader frequency range than Debye type processes with a discrete relaxation time (e. g. Eq. (20)). An example of a spectrum is shown in Fig. 20. Also

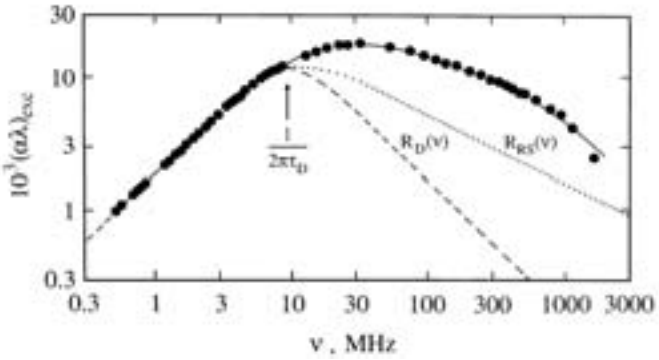


Figure 20. Ultrasonic excess attenuation spectrum of a mixture of water and 2-(2-butoxyethoxy)ethanol (C_4E_2) at 25°C [88]. The mole fraction of C_4E_2 is $x = 0.04$. Dashed and dotted lines are graphs of a Debye relaxation spectral term with discrete relaxation time τ_D [34] and of a Romanov-Solov'ev term [91], respectively. The full curve represents the unifying model of noncritical concentration fluctuations ([89], Eq. (53)).

given in that diagram is the graph of a relaxation term $R_{\mathrm{RS}}(\nu)$ as resulting from the Romanov-Solov'ev theory of noncritical concentration fluctuations [91–93]. This term also represents the experimental data only insufficiently. Several extensions to the Romanov-Solov'ev theory have been made [87,89,94,95]. Here we sketch only the last unifying model [89] because it combines the relevant aspects of all previous theories.

In the unifying model changes in the local composition of the binary liquids are assumed to occur along two possible pathways, one of which is an elementary chemical reaction with discrete relaxation time τ_0 and the other one a diffusion process with mutual diffusion coefficient D. The time behaviour of the fluctuations is then controlled by the differential equation

$$\frac{\partial \Phi(\boldsymbol{r},t)}{\partial t} = \left(D\nabla^2 - \frac{1}{\tau_0}\right)\Phi(\boldsymbol{r},t) \tag{47}$$

with $\Phi(\boldsymbol{r},t)$ denoting the autocorrelation function of the order parameter, namely the deviation of the local concentration from the mean. Spatial Fourier transformation yields

$$\hat{\Phi}(\boldsymbol{q},t) = \int_{\boldsymbol{r}} \Phi(\boldsymbol{r},t)\exp(\imath\boldsymbol{r}\boldsymbol{q})\,d\boldsymbol{r} \tag{48}$$

with wave vector $\boldsymbol{q}$. In the q space the simpler differential equation

$$\frac{\partial \hat{\Phi}(\boldsymbol{q},t)}{\partial t} = -\frac{1}{\tau_g}\hat{\Phi}(\boldsymbol{q},t) \tag{49}$$

follows with

$$\tau_g^{-1} = Dq^2 + \tau_0^{-1}\,. \tag{50}$$

Here $q = |\boldsymbol{q}|$. For isotropic liquids the correlations in the local fluctuations will depend on the distance $r = |\boldsymbol{r}|$ only. Eq. (49) may thus be solved by exponentials with decay time τ_g for each Fourier component. Hence

$$\hat{\Phi}(q,t) = \hat{f}(q)\exp(-t/\tau_g)\,. \tag{51}$$

The weight function $\hat{f}(q)$ is the Fourier transformation of the correlations in space $\Phi(r,0)$ of the concentration fluctuations at time $t = 0$. The different theories of noncritical concentration fluctuations differ from one another by the weight function. The unifying model uses the function

$$\hat{f}(q) = \left(1 + 0.164(q\xi) + 0.25(q\xi)^2\right)^{-2}\,, \tag{52}$$

assuming the long-range correlations to follow Ornstein-Zernike behaviour, whereas short-range correlations are considered by a nearly exponential decay at $r < \xi$. The special choice of the weight function leads to a contribution from concentration fluctuations to the relaxation spectral function

$$R_{\mathrm{um}}(\nu) = Q\int_0^\infty \hat{f}(q)\frac{\omega\tau_g}{1+\omega^2\tau_g^2}q^2dq \tag{53}$$

describing the response of the liquid to compressional waves without artifical limits in the integration procedure. The amplitude factor Q is assumed a sum of the Romanov-Solov'ev factor Q_{RS} [91], given by the relation

$$Q_{\mathrm{RS}} = \frac{\varrho c^2 k_B T}{8\pi}\frac{V^2}{g''^2}\left(\frac{v''}{V} - \Lambda\frac{h''}{C_p}\right)\,, \tag{54}$$

and the Montrose-Litovitz contribution Q_{ML} [94], resulting from a shear viscosity relaxation with identical frequency characteristics. In Eq. (54) k_B is Boltzmann's constant and the double-primed quantities

$$g'' = \frac{\partial^2 G_0}{\partial\bar{x}_2^2}\ ,\ v'' = \frac{\partial^2 V_0}{\partial\bar{x}_2^2}\ ,\ h'' = \frac{\partial^2 H_0}{\partial\bar{x}_2^2} \tag{55}$$

are the second derivatives of the Gibbs free enthalpy, the molar volume, and the molar enthalpy, respectively, without contributions from the concentration fluctuations. Parameter $\bar{x}_2$ denotes the equilibrium mole fraction of the dispersed phase. For an example Fig. 21 shows the Q_{RS}-versus-$\bar{x}_2$ relation, as following from Eqs. (54) and (55), along with experimental Q data. Even though it is difficult to derive reliable second derivative values from experimental G_0, V_0, and H_0 data the agreement between theory and experiment is satisfactory. This is a remarkable result as the amplitude factor displays a quite unusual dependence upon $\bar{x}_2$.

Using diffusion coefficients D as resulting from an analysis of experimental spectra in terms of relaxation function $R_{\mathrm{um}}(\nu)$ and taking shear viscosity η_s from measurements the fluctuation correlation length ξ can be derived from the Kawasaki-Ferrell relation [84,96]

$$\xi = \frac{k_B T}{6\pi\eta_s D}\,. \tag{56}$$

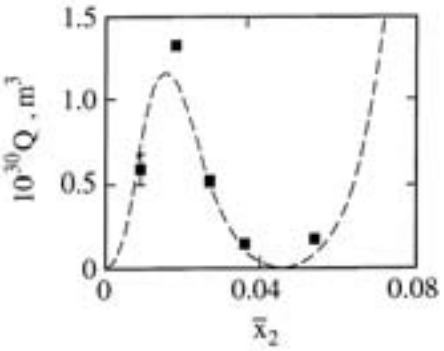


Figure 21. Amplitude factor of the relaxation spectral function $R_{um}(\nu)$ (Eq. (53)) for aqueous solutions of tetra-*n*-propylammonium bromide at 25°C, displayed as a function of equilibrium mole fraction $\bar{x}_2$ of salt [87,89]. The dashed line is the graph of the Romanov-Solov'ev ampitude factor (Eqs. (54,55)).

For three series of solutes the maximum ξ values of aqueous systems are displayed in Fig. 22 as a function of the number n of alkyl groups per molecule of solute. Within the series of unbranched molecules the maximum correlation length $\xi_{\max}$ increases significantly with length of the hydrophobic group of solute. Obviously, the nature of the hydrophilic group is of low significance for $\xi_{\max}$. In correspondence to the formation of micelles, the larger the hydrophobic part of the nonaqueous constituent

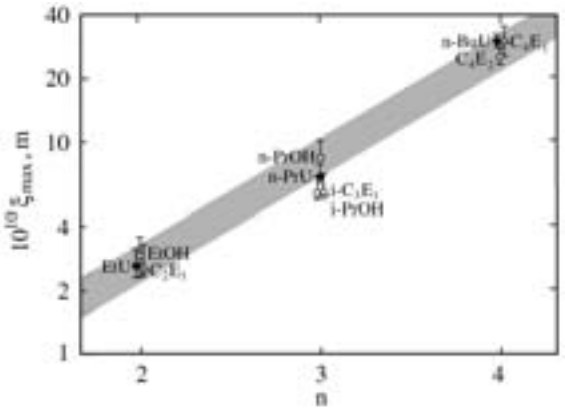


Figure 22. Maximum value $\xi_{\max}$ in the concentration dependence of a binary system, displayed as a function of the number n of alkyl groups per organic molecule for some aqueous solutions of urea derivatives (•), monohydric alcohols (○), and poly(ethylene glycol) monoalkyl ethers (△) at 25°C [89]. EtU, ethylurea; *n*-PrU, *n*-propylurea; *n*-BuU, *n*-butylurea; EtOH, ethanol; *n*-PrOH, *n*-propanol; i-PrOH, 2-propanol; C_2E_1, 2-butoxyethanol; i-C_3E_1, isopropoxyethanol; C_4E_1, 2-butoxyethanol; C_4E_2, 2-(2-butoxyethoxy)ethanol.

the stronger the tendency to form clusters in order to avoid unfavourable interactions with water. The effect of cluster formation is noticeably smaller for solutes with branched hydrophobic parts. Almost no fluctuations of local concentration exist for the tetramethylurea/water system ($\xi_{\max} = 2 \cdot 10^{-10}$ m [90]), whereas $\xi_{\max} = 30 \cdot 10^{-10}$ m for aqueous solutions of *n*-butylurea (Fig. 22).

4.2 Critical demixing

The anomalies in the thermodynamic and transport properties, which are induced by long wavelength fluctuations in the order parameter, associated with the phase transition near a critical point, have attracted much interest from both a theoretical and an experimental point of view. Considerable attention has been directed towards ultrasonic attenuation spectrometry as the method allows to verify or disprove the dynamic scaling hypothesis, particularly if combined with quasielastic light scattering and shear viscosity measurements. Various theoretical models for the description of ultrasonic spectra of critical systems exist, the most prominent are the dynamic scaling theory [97–100], the mode coupling theory [101–104], and the more recent intuitive theory proceeding from a description of the bulk vicosity near the critical point [105,106].

The dynamic scaling model predicts the critical contribution

$$\alpha_\lambda^c = \alpha_\lambda - \alpha_\lambda^b \tag{57}$$

to the total attenuation per wavelength, $\alpha_\lambda = \alpha\lambda$, to be given by

$$\alpha_\lambda^c = cA(T)F(\Omega)\,. \tag{58}$$

In this relation A is an amplitude factor, only weakly depending on frequency, and $F(\Omega)$ is the scaling function with reduced frequency

$$\Omega = \frac{2\pi\nu}{\Gamma(\epsilon)}\,. \tag{59}$$

The relaxation rate of order parameter fluctuations, $\Gamma(\epsilon) = \tau_\xi^{-1}$, is assumed to follow a power law

$$\Gamma(\epsilon) = \Gamma_0 \epsilon^{-Z_0\tilde{\nu}} \tag{60}$$

with the universal dynamical critical exponent Z_0 and the critical exponent $\tilde{\nu}$ of the fluctuation correlation length mentioned before.

In correspondence with an empirical form of the dynamic scaling model [99] the scaling functions of the Bhattacharjee-Ferrell (BF), Folk-Moser (FM), and Onuki (On) models can be favourably represented by the relation [107]

$$F_x(\Omega) = \left[1 + 0.414\left(\frac{\Omega_{1/2}^x}{\Omega}\right)^{n_x}\right]^{-2} \tag{61}$$

with $\Omega_{1/2}^x$ (x = BF, FM, On) denoting the scaled half-attenuation frequency of F_x and n_x an exponent that controls the slope $S_x(\Omega = \Omega_{1/2}^x) = dF_x(\Omega)/d\ln(\Omega)|_{\Omega_{1/2}}$ of

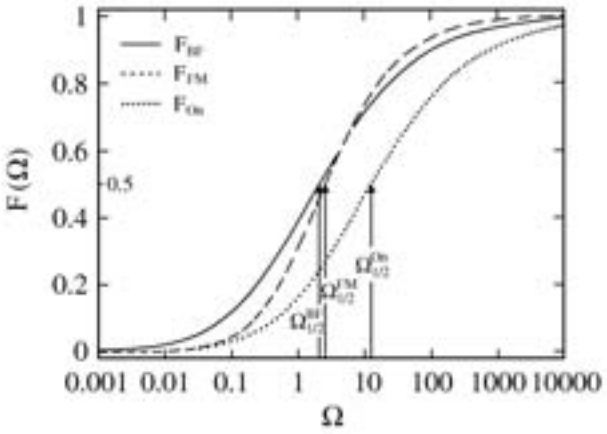


Figure 23. Graphs of the scaling functions (Eq. (61)) of the Bhattacharjee-Ferrell (BF), Folk-Moser (FM), and Onuki (On) theory. Arrows indicate the half-attenuation frequencies [107].

the scaling function at its half-attenuation frequency. Graphs of the scaling functions from the three theoretical models are shown in Fig. 23. The corresponding parameter values are $\Omega_{1/2}^{\mathrm{BF}} = 2.1$, $\Omega_{1/2}^{\mathrm{FM}} = 3.1$, and $\Omega_{1/2}^{\mathrm{On}} = 6.2$, as well as $n_{\mathrm{BF}} = n_{\mathrm{On}} = 0.5$ and $n_{\mathrm{FM}} = 0.635$ [107].

Recently a variety of binary mixtures has been studied in order to find out which of the theoretical scaling functions fits best to the experimental findings. For these investigations systems with an as simple as possible background part in the attenuation-per-wavelength spectra have been chosen because interferences of the critical dynamics with noncritical processes are largely avoided thereby [108–114]. The scaling function has been directly calculated from the experimental data using the relation [99]

$$F(\Omega) = \alpha_\lambda^c(\nu, T)/\alpha_\lambda^c(\nu, T_c)\,. \tag{62}$$

The relaxation rate Γ, required for the determination of $F(\Omega)$, has been obtained from dynamic light scattering experiments, yielding the mutual diffusion coefficient, and shear viscosity measurements. Considering effects of the crossover from Ising to mean field behaviour the shear viscosity η_s has been evaluated in terms of the dynamic scaling theory [115], which predicts

$$\eta_s(\epsilon) = \eta_b(\epsilon)\exp(Z_\eta H)\,, \tag{63}$$

with the universal critical exponent Z_η of the shear viscosity ($Z_\eta = 0.065$ [116,117]), with the background viscosity

$$\eta_b(\epsilon) = A_\eta \exp\left[\frac{B_\eta}{T - T_\eta}\right]\,, \tag{64}$$

and with the crossover function $H = H(\xi, q_c, q_D)$, depending upon the fluctuation correlation length ξ and on two cutoff wave numbers q_c and q_D. The analytical form

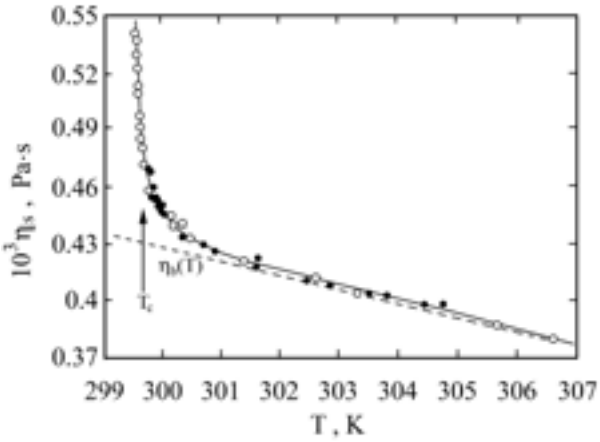


Figure 24. Shear viscosity η_s of the nitroethane-3-methylpentane mixture of critical composition displayed versus temperature T [111]. Circles show previous data [121]. The full curve is the graph of the theoretical η_s relation (Eq. (63)), the dashed line represents the noncritical background contribution η_s (Eq. (64)).

of H is given elsewhere [115,116]. Use of the more recent H' crossover function [118] does not noticeably change the data obtained from the evaluation procedure. The individual parameters A_η, B_η, and T_η in Eq. (64) are characteristic of the system under consideration.

The mutual diffusion coefficient

$$D = D_{\mathrm{KF}} \left[\left(1 + b^2 x^2\right)^{Z_\eta/2} R\,\Omega_{\mathrm{K}}(x) + \frac{3\pi\eta_s}{16\eta_b} \frac{1+x^2}{\xi} \left(\tilde{q}_c^{-1} - q_D^{-1}\right) \right] \tag{65}$$

depends also on parameters ξ, q_c, and q_D. Here $x = q\xi$ with q denoting the amount of the wave vector selected by scattering geometry in the dynamic light scattering measurements. Furtheron, $b = 0.55$, $R = 1.03$, $\tilde{q}_c^{-1} = q_c^{-1} + (2q_D)^{-1}$, and

$$\Omega_{\mathrm{K}}(x) = \frac{3}{4x^2} \left[1 + x^2 + (x^3 - x^{-1)}) \arctan x \right] \tag{66}$$

is the Kawasaki function [119]. Using Eq. (45) with $\tilde{\nu} = 0.63$ the unknown parameters ξ_0, q_c, q_D, A_η, B_η, and T_η follow from a simultaneous regression analysis of the shear viscosity and diffusion coefficient data. An example of shear viscosity data and their representation by Eq. (63) is presented in Fig. 24. Figure 25 shows the relaxation rates following from the shear viscosity and dynamic light scattering results. This diagram also indicates the adequacy of the power law (Eq. (60)). For the same binary system the scaling function data resulting from Eq. (62) are shown in Fig. 26, indicating that the Bhattacharjee-Ferrell scaling function represents the experimental data within their limits of experimental errors. The same result has been found for the other systems investigated. The parameter values for these systems vary between $\xi_0 = 0.145\,\mathrm{nm}$ and $\Gamma_0 = 187 \cdot 10^9\,\mathrm{s}^{-1}$, n-pentanol-nitromethane [109], and

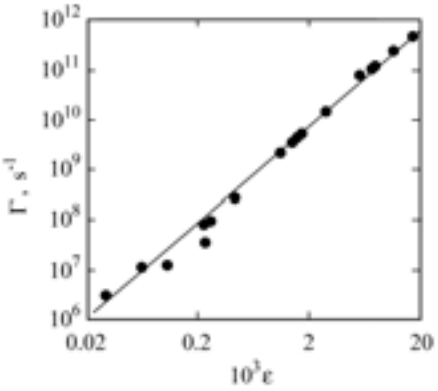


Figure 25. Relaxation rate Γ of concentration fluctuations versus reduced temperature ϵ for the nitroethane-3-methylpentane mixture of critical composition [111]. The line in the graph of the power law (Eq. (60)) with universal exponent $Z_0\tilde{\nu} = 3.065$ [6] and with individual amplitude $\Gamma_0 = 125 \cdot 10^9\,\mathrm{s}^{-1}$ [111].

$\xi_0 = 0.37\,\mathrm{nm}$ as well as $\Gamma_0 = 6.4 \cdot 10^9\,\mathrm{s}^{-1}$, ethanol-dodecane [113]. The inset in Fig. 26 presents the scaled half-attenuation frequency which, when treated as an unknown parameter, follows as

$$\Omega_{1/2}^{\mathrm{BF}} = \Omega \left[\frac{1}{0.414} \left(F_{\mathrm{BF}}^{-1/2}(\Omega) - 1 \right) \right]^2 . \tag{67}$$

The small scatter of the experimental data around the theoretically predicted value 2.1 emphasizes the appropriateness of the Bhattacharjee-Ferrell dynamic scaling model.

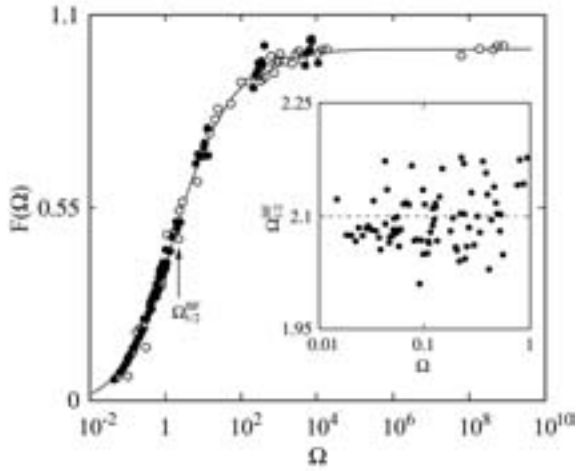


Figure 26. Scaling function data according to Eq. (62): • [111], ○ [121]. The full line is the graph of the empirical Bhattarcharjee-Ferrell scaling function (Eq. (61) with $x = \mathrm{BF}$). For small Ω values the half-attenuation frequency data as calculated according to Eq. (67) are displayed in the inset. The dashed line indicates the theoretically predicted $\Omega_{1/2}^{\mathrm{BF}} = 2.1$ [99].

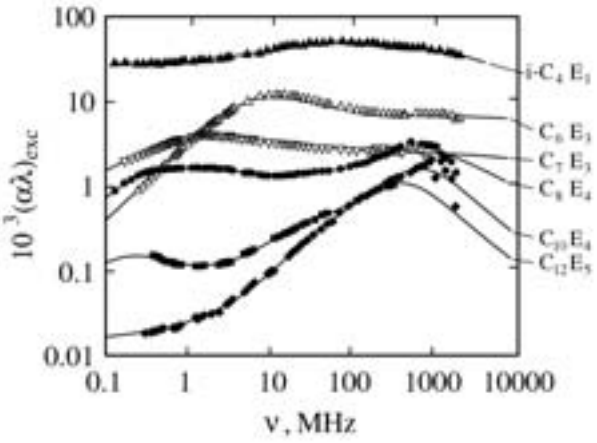


Figure 27. Ultrasonic excess attenuation spectra of some C_iE_j-water mixtures: $i - C_4E_1$, isobutoxyethanol, $Y = 0.33$, 25°C; C_6E_3, triethyleneglycol monhexyl ether, $Y = 0.08$, 21°C; C_8E_4, tetraethyleneglycol monooctyl ether, $Y = 0.071$, 25°C; $C_{10}E_4$, tetraethyleneglycol monodecyl ether, $Y = 0.02$, 16°C; $C_{12}E_5$, pentaethyleneglycol monododecyl ether, $Y = 0.015$, 25°C. Y denotes the mass fraction of nonionic surfactant. The lines are graphs of the spectral functions that analytically represent the spectra [16].

4.3 Concentration fluctuations and micelle formation

Systems displaying critical dynamics and, simultaneously, noncritical elementary reactions, such as conformational isomerisations, associations, and protolysis/hydrolysis [122–127], as well as micelle formation/decay kinetics [15–17,81,128–131], have been the focus of special interest recently. Most studies aimed particularly at the slowing down of chemical reactions near a critical point [132,133]. Here we mention only some results for nonionic surfactant solutions which exhibit both, a critical micelle concentration and a critical demixing point. Such systems attract attention due to a second critical phenomenon, besides the fluctuations in the local concentration, and due to the presence of a second characteristic length, the diameter of micelles, in addition to the fluctuation correlation length.

Challenging aspects for complex liquid studies offer poly(ethylene glycol) monoalkyl ethers ($CH_3(CH_2)_{i-1}(OCH_2CH_2)_jOH$; $i = 1, 2, \ldots$; $j = 0, 1, \ldots$, abbreviated C_iE_j), as most of them form micelles and microheterogeneous phases as well. The variability in the number of nonassociating hydrophobic (C_i) and hydrogen bonding hydrophilic (E_j) group enables the kinetics of micelle formation and the critical dynamics to be investigated for homologous series of surfactants. Figure 27 indicates the noticeable variety in the shape of the sonic spectra of nonionic surfactant systems and also the broad range of maximum excess attenuation values. For the system triethylene glycol monoheptyl ether-water two ultrasonic attenuation terms contribute to the spectrum (Fig. 28), one representing the Bhattacharjee-Ferrell critical contribution. The other one can be well represented by the Hill relaxation function (Eq. (37)) as characteristic for the fast monomer exchange of micellar solutions [17]. Though the experimental

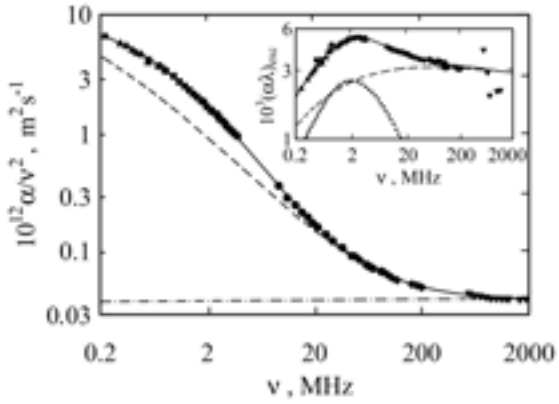


Figure 28. Broadband ultrasonic attenuation spectrum of the triethyleneglycol monoheptyl ether-water mixture of critical composition (C_7E_3-H_2O, $Y = 0.1$) at 15°C displayed in both common formats. Full lines are graphs of the relaxation function representing the spectrum, dashed lines represent the Bhattacharjee-Ferrell critical term. The dashed-dotted line shows the high-frequency contribution to the α/ν^2 values and the dotted line indicates the Hill type relaxation term reflecting the fast monomer exchange of the surfactant system.

spectra can be analyzed in terms of a linear superposition of both relaxation terms, a coupling between the fluctuations in the local concentration and the kinetics of micelle formation seems to exist. An indication for interferences of the critical dynamics with the micelle kinetics is the small value $\Omega_{1/2}^{\mathrm{BF}} = 1$ for the half-attenuation frequency in the scaling function, which deviates considerably from the value $\Omega_{1/2}^{\mathrm{BF}} = 2.1$ as characteristic for simpler critical mixtures (Sect. 4.2). Such interferences are also revealed by the remarkable correlation between the maximum attenuation-per-wavelength value $(\alpha\lambda)_{\mathrm{exc}}^{\mathrm{max}}$ in a series of C_iE_j/H_2O mixtures and the critical micelle concentration of that system (Fig. 29). This correlation, which exists for a broad range of cmc values, has stimulated a fluctuation controlled monomer exchange model [15]. Apart from a critical demixing point, no noticeable fluctuations in the local concentrations of micelles and monomers exist. Hence the micelle kinetics is governed by the monomer exchange as predicted by the Teubner-Kahlweit-Aniansson-Wall model or its extended version described before (Sect. 3.2). Approaching the critical demixing point, diffusion controlled local fluctuations in the spatial distribution of micelles, accompanied by fluctuations in the distribution of monomers, act a noticeable effect on the molecular dynamics. Due to the law of mass action the monomer exchange kinetics, via the rate equations, is largely governed by the fluctuations in the local concentration of micelles. The coupling between concentration fluctuations and the micelle formation/decay process suggests the linear superposition of relaxation terms in the ultrasonic spectra as a first approximation. Strictly the theoretical model of sonic attenuation should combine both molecular mechanisms in a comprehensive treatment.

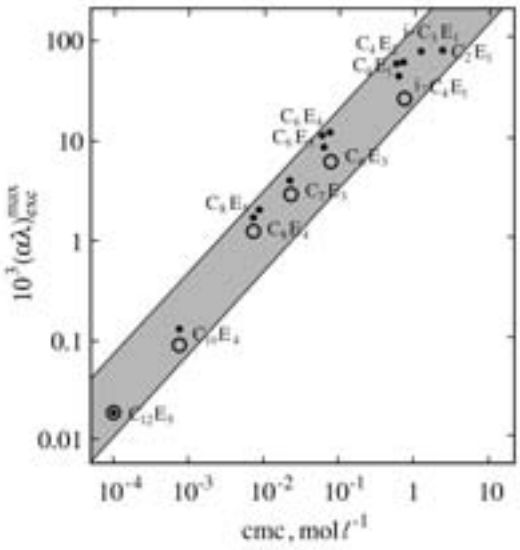


Figure 29. Maximum values in the excess attenuation spectra (•) and the critical contribution to the spectra (○) displayed as a function of critical micelle concentration for a variety of $C_iE_j - H_2O$ mixtures at 25°C (C_7E_3 at 22.5°C; $C_{10}E_4$ at 18°C [16]).

5 Conclusions

Liquids are characterized by a rich variety of rapidly fluctuating structures, including short living dimers as well as complexes formed from micelles, being aggregates themselves. The diversity in the molecular dynamics resulting thereby can be favourably studied by acoustical spectrometry. Compressional waves couple to nearly all relevant molecular mechanisms, covering a broad range of frequencies and thus relaxation times. Though naturally rather nonspecific, variations of the temperature of the liquid and of the concentration and chemical composition of constituents enable definite results on thermodynamic and tranport parameters, which are difficult to obtain otherwise. Specific insights into even complex dynamics is enabled thereby, especially if additional experimental methods are used for complementary information.

References

[1] I. Ohmine and H. Tanaka, 'Fluctuation, Relaxations, and Hydration in Liquid Water. Hydrogen-Bond Rearrangement Dynamics', Chem. Rev. **93**, 2545 (1993).
[2] U. Kaatze, R. Behrends, and R. Pottel, 'Hydrogen Network Fluctuations and Dielectric Spectrometry of Liquids', J. Non-Crystalline Solids **305**, 19 (2002).
[3] U. Kaatze, in *Electromagnetic Aquametry, Electromagnetic Wave Interaction with Water and Moist Substances*, edited by K. Kupfer (Springer, Berlin, 2005).
[4] A. M. Belloc, *Membranes, Microemulsions and Monolayers* (Springer, Berlin, 1994).
[5] D. F. Evans and H. Wennerström, *The Colloidal Domain. Where Physics, Chemistry, and Biology Meet* (Wiley-VCH, New York, 1999).

[6] M. A. Anisimov, *Critical Phenomena in Liquids and Liquid Crystals* (Gordon and Breach, Philadelphia, 1991).
[7] C. Domb, *The Critical Point: A Historical Introduction to the Modern Theory of Critical Phenomena* (Taylor and Francis, London, 1996).
[8] A. Onuki, *Phase Transition Dynamics* (Cambridge University Press, Cambridge, 2002).
[9] L. M. Brown, *Renormalization* (Springer, Berlin, 1993).
[10] M. E. Fisher, 'Renormalization Group Theory: Its Basis and Formulation in Statistical Physics', Rev. Mod. Phys. **70**, 653 (1998).
[11] H. E. Stanley, 'Scaling, Universility, and Renormalization: Three Pillars of Modern Critical Phenomena', Rev. Mod. Phys. **71**, S358 (1999).
[12] J. Israelachvili, *Intermolecular and Surface Forces* (Academic, New York, 1992).
[13] W. M. Gelbart, A. Ben-Shaul, and D. Roux (Eds.), *Micelles, Membranes, Microemulsions, and Monolayers* (Springer, Berlin, 1994).
[14] B. Jonsson, B. Lindman, K. Holmberg, and B. Kronberg, *Surfactants and Polymers in Aqueous Solution* (Wiley, Chichester, 1998).
[15] T. Telgmann and U. Kaatze, 'Monomer Exchange and Concentration Fluctuations in Poly(ethylene glycol)Monoalkyl Ether/Water Mixtures. Toward a Uniform Description of Acoustical Spectra', Langmuir **18**, 3068 (2002).
[16] E. Hanke, T. Telgmann, and U. Kaatze, 'Monomer Exchange Kinetics, Dynamics of Concentration Fluctuations, and Chain Isomerization of Nonionic Surfactant/Water Systems. Evidence From Broadband Ultrasonic Spectra', Tenside Surf. Det. **41**, 23 (2005).
[17] J. Haller, R. Behrends, and U. Kaatze, 'Critical Fluctuations of Micellar Triethylene Glycol Monoheptyl Ether-Water System', J. Chem. Phys. **124**, 124910 (2006).
[18] G. Kurtze and K. Tamm, 'Measurements of Sound Absorption in Water and in Aqueous Solutions of Electrolytes', Acustica **3**, 33 (1953).
[19] M. Eigen and K. Tamm, 'Schallabsorption in Elektrolytlösungen als Folge chemischer Relaxation I. Relaxationstheorie der mehrstufigen Dissoziation', Z. Elektrochem. Ber. Bunsenges. Phys. Chem. **66**, 93 (1962).
[20] M. Eigen and K. Tamm, 'Schallabsorption in Elektrolytlösungen als Folge chemischer Relaxation II. Meßergebnisse und Relaxationsmechanismen für 2-2-wertige Elektrolyte', Z. Elektrochem. Ber. Bunsenges. Phys. Chem. **66**, 107 (1962).
[21] K. G. Plaß, 'Relaxation in organischen Flüssigkeiten', Acustica **19**, 236 (1967/68).
[22] F. Bader and K. G. Plaß, 'Sound Absorption in Liquid Acetic Acid in the Frequency Range between 0.3 and 1.5 GHz', Ber. Bunsenges. Physik. Chem. **75**, 553 (1971).
[23] F. Eggers, 'Eine Resonanzmethode zur Bestimmung von Schallgeschwindigkeit und Dämpfung an geringen Flüssigkeitsmengen', Acustica **19**, 323 (1967/68).
[24] A. B. Bhatia, *Ultrasonic Absorption* (Oxford University Press, Oxford, 1967).
[25] E. Hanke and U. Kaatze, 'Structural Aspects and Molecular Dynamics of Liquid Polymers, Including Mixtures with Water', J.Phys. IV France **129**, 29 (2005).
[26] E. Hanke, *Struktureigenschaften und molekulare Dynamik flüssiger Ethylenglykol-Oligomere und ihrer Mischungen mit Wasser*, Dissertation, Georg-August-Universität Göttingen (2007).
[27] P. E. Rouse Jr., 'A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers', J. Chem. Phys. **21**, 1272 (1953).
[28] B. Zimm, 'Dynamics of Polymer Molecules in Dilute Solutions: Viscoelasticity, Flow Birefringence and Dielectric Loss', J. Chem. Phys. **24**, 269 (1956).
[29] A. V. Tobolsky and J. J. Aklonis, 'A Molecular Theory for Viscoelastic Behavior of Amorphous Polymers', J. Chem. Phys. **68**, 1970 (1964).

[30] A. V. Tobolsky and D. B. DuPré, 'Macromolecular Relaxation in the Damped Torsional Oscillator and Statistical Segment Models', Adv. Polym. Sci. **6**, 103 (1969).
[31] R. Behrends and U. Kaatze, 'Structural Isomerizaion and Molecular Motions of Liquid n-Alkanes. Ultrasonic and High-Frequency Shear Viscosity Relaxation', J. Phys. Chem. A **104**, 3269 (2000).
[32] R. Behrends and U. Kaatze, 'Hydrogen Bonding and Chain Conformational Isomerization of Alcohols Probed by Ultrasonic Absorption and Shear Impedance Spectrometry', J. Phys. Chem. A **105**, 5829 (2001).
[33] R. Behrends and U. Kaatze, 'A High Frequency Shear Wave Impedance Spectrometer for Low Viscosity Liquids', Meas. Sci. Technol. **12**, 519 (2001).
[34] P. Debye, *Polar Molecules* (Chemical Catalog Co., New York, 1929).
[35] R. Polacek and U. Kaatze, 'A Small Volume Spherical Resonator Method for the Acoustical Spctrometry of Liquids Down to Audio Frequencies', Meas. Sci. Technol. **12**, 1 (2001).
[36] U. Kaatze, B. Wehrmann, and R. Pottel, 'Acoustical Absorption Spectroscopy of Liquids Between 0.15 and 3000 MHz: I. High Resolution Ultrasonic Resonator Method', J. Phys. E: Sci. Instrum. **20**, 1025 (1987).
[37] F. Eggers, U. Kaatze, K. H. Richmann, and T. Telgmann, 'New Plano-Concave Ultrasonic Resonator Cells for Absorption and Velocity Measurements in Liquids Below 1 MHz', Meas. Sci. Technol. **5**, 1131 (1994).
[38] R. Behrends, F. Eggers, U. Kaatze, and T. Telgmann, T., 'Ultrasonic Spectrometry of Liquids Below 1 MHz. Biconcave Resonator Cell with Adjustable Radius of Curvature', Ultrasonics **34**, 59 (1996).
[39] R. Polacek and U. Kaatze, 'A High-Q-Easy-to-Handle Biconcave Resonator for Acoustic Spectrometry of Liquids', Meas. Sci. Techn. **14**, 1068 (2003).
[40] U. Kaatze, K. Lautscham, and M. Brai, 'Acoustical Absorption Spectroscopy of Liquids Between 0.15 and 3000 MHz: II. Ultrasonic Pulse Transmission Methods', J. Phys. E: Sci. Instrum. **21**, 98 (1988).
[41] U. Kaatze, V. Kühnel, K. Menzel, and S. Schwerdtfeger, 'Ultrasonic Spectroscopy of Liquids. Extending the Frequency Range of the Variable Sample Length Pulse Technique', Meas. Sci. Techn. **4**, 1257 (1993).
[42] U. Kaatze, V. Kühnel, and G. Weiss, 'Variable Pathlength Cells for Precise Hypersonic Spectrometry of Liquids up to 5 GHz', Ultrasonics **34**, 51 (1994).
[43] F. Eggers and U. Kaatze, 'Broad-Band Ultrasonic Measurement Techniques for Liquids', Meas. Sci. Technol. **7**, 1 (1996).
[44] R. Polacek, J. Stenger, and U. Kaatze, 'Chair-Chair Conformational Flexibility, Pseudorotation, and Exocyclic Group Isomerization of Monosaccharides in Water', J. Chem. Phys. **116**, 2973 (2002).
[45] U. Kaatze and B. Wehrmann, 'Broadband Ultrsonic Spectroscopy of Aqueous Solutions of Zinc(II)Chloride :I. Kinetics of Complexation', Z. Phys. Chem (Munich) **177**, 9 (1992).
[46] E. Baucke, R. Behrends, K. Fuchs, R. Hagen, and U. Kaatze, 'Kinetics of Ca^{2+} Complexation with Some Carbohydrates in Aqueous Solutions', J. Chem. Phys. **120**, 8118 (2004).
[47] M. Cowman, F. Eggers, E. M. Eyring, D. Horoszewski, U. Kaatze, R. Kreitner, S. Petrucci, M. Klöppel-Riech, and J. Stenger, 'Microsecond to Subnanosecond Molecular Relaxation Dynamics of the Interaction of Ca^{2+} with Some Carbohydrates in Aqueous Solutions', J. Phys. Chem. B **103**, 239 (1999).
[48] R. Behrends and U. Kaatze, 'Molecular Dynamics and Conformational Kinetics of Mono- and Disaccharides in Aqueous Solution', ChemPhysChem **6**, 1133 (2005).

[49] M. Eigen and R. Winkler, in *Neuroscience: Second Study Program*, edited by F. O. Schmitt (Rockefeller University Press, New York, 1970).
[50] C. S. Nascimento Jr., H. F. Dos Santos, and W. B. De Almeida, 'Theoretical Study of the Formation of the α-Cyclodextrin Hexahydrate', Chem. Phys. Lett. **397**, 422 (2004).
[51] S. Nishikawa, T. Fukahori, and K. Ishikawa, 'Ultrasonic Relaxations in Aqueous Solutions of Propionic Acid in the Presence and Absence of β-Cyclodextrin', J. Phys. Chem. A **106**, 3029 (2002).
[52] T. Fukahori, T. Ugawa, and S. Nishikawa, 'Molecular Recognition Kinetics of Leucine and Glycol-Leucine by β-Cyclodextrin in Aqueous Solution in Terms of Ultrasonic Relaxation', J. Phys. Chem. A **106**, 9442 (2002).
[53] T. Fukahori, S. Nishikawa, and K. Yamaguchi, 'Ultrasonic Relaxation Due to Inclusion Complex of Amino Acid by β-Cyclodextrin in Aqueous Solution', J. Acoust. Soc. Am. **115**, 2325 (2004).
[54] T. Fukahori, S. Nishikawa, and K. Yamaguchi, 'Kinetics on Isomeric Alcohols Recognition by α- and β-Cyclodextrins Using Ultrasonic Relaxation Method', Bull. Chem. Soc. Jpn. **77**, 2193 (2004).
[55] Y. Yu, C. Chipot, W. Cai, and X. Shao, 'Molecular Dynamics Study of the Inclusion of Cholesterol into Cyclodextrins', J. Phys. Chem. B **110**, 6372 (2006).
[56] J. W. Minns and A. Khan, 'α-Cyclodextrin-I_3^- Host-Guest Complex in Aqueous Solution: Theoretical and Experimental Studies', J. Phys. Chem. A **106**, 6421 (2002).
[57] S. Nishikawa, K. Yamagushi, and T. Fukahori, 'Ultrasonic Relaxation Due to Complexation Reaction Between β-Cyclodextrin and Alkylammonium Ions', J. Phys. Chem. A **107**, 6415 (2003).
[58] J. C. Papaioannou and T. C. Ghikas, 'Dielectric Relaxation of α-Cyclodextrin-Polyiodide Complexes (α-Cyclodextrin)$_2 \cdot BaI_2 \cdot I_2 \cdot 8H_2O$', Mol. Phys. **101**, 2601 (2003).
[59] K. Yamagushi, T. Fukahori, and S. Nishikawa, 'Dynamic Interaction Between Alkylammonium Ions and β-Cyclodextrin by Means of Ultrasonic Relaxation', J. Phys. Chem. A **109**, 40 (2005).
[60] J. Haller, P. Miecznik, and U. Kaatze, 'Ultrasonic Attenuation Spectrometry Study of α-Cyclodextrin + KI Complexation in Water', Chem. Phys. Lett. **429**, 97 (2006).
[61] A. Hazekamp and R. Verpoorte, 'Structure Elucidation of Tetrahydrocannabinal Complex with Randomly Methylated β-Cyclodextrin', Eur. J. Pharmac. Sci. **29**, 340 (2006).
[62] T. Fukahori, M. Kondo, and S. Nishikawa, 'Dynamic Study of Interaction Between β-Cyclodextrin and Aspirin by Ultrasonic Relaxation Method', J. Phys. Chem. B **110**, 4487 (2006).
[63] L. De Maeyer, C. Trachimow, and U. Kaatze, 'Entropy-Driven Micellar Aggregation', J. Phys. Chem. B **102**, 8480 (1998).
[64] R. Polacek, V. A. Buckin, F. Eggers, and U. Kaatze, 'Kinetics of Base-Stacking Interactions and Proton Exchange of 6-Methylpurine Aqueous Solutions', J. Phys. Chem. A **108**, 1867 (2004).
[65] E. A. G. Aniansson and S. N. Wall, 'On the Kinetics of Step-Wise Micelle Association', J. Phys. Chem. **78**, 1024 (1974).
[66] E. A. G. Aniansson, 'The Mean Lifetime of a Micelle', Prog. Colloid Polym. Sci. **70**, 2 (1985).
[67] M. Teubner, 'Theory of Ultrasonic Absorption in Micellar Solutions', J. Phys. Chem. **83**, 2917 (1979).
[68] M. Kahlweit and M. Teubner, 'On the Kinetics of Micellation in Aqueous Solutions',

Adv. Colloid Interface Sci. **13**, 1 (1980).
[69] H. Strehlow, *Rapid Reactions in Solution* (VCH, Weinheim, 1992).
[70] U. Kaatze, K. Lautscham, and W. Berger, 'Ultra- and Hypersonic Absorption and Molecular Relaxation in Aqueous Solutions of Anionic and Cationic Micelles', Z. Phys. Chem. (Munich) **159**, 161 (1988).
[71] R. Polacek and U. Kaatze, 'Monomer Exchange Kinetics, Radial Diffusion, and Hydrocarbon Chain Isomerization of Sodium Dodecylsulfate Micelles in Water', J. Phys. Chem. B **111**, 1625 (2007).
[72] E. Lessner, M. Teubner, and M. Kahlweit, 'Relaxation Experiments in Aqueous Solutions of Ionic Micelles. 1. Theory and Experiments on the System H_2O-Sodium Tetradecyl Sulfate-$NaClO_4$', J. Phys. Chem. **85**, 1529 (1981).
[73] D. G. Hall and E. Wyn-Jones, 'Chemical Relaxation Spectrometry in Aqueous Surfactant Solutions', J. Mol. Liq. **32**, 63 (1986).
[74] T. Telgmann and U. Kaatze, 'On the Kinetics of the Formation of Small Micelles. 1. Broadband Ultrasonic Absorption Spectrometry', J. Phys. Chem. B **101**, 7758 (1997).
[75] Telgmann, T. and Kaatze, U., On the Kinetics of the Formation of Small Micelles. 2. Extension of the Model of Stepwise Association, J.Phys.Chem.B, 1997, 101, 7766
[76] R. Polacek, *Breitbandige Ultraschallabsorptionsspektroskopie an wässrigen ionischen Tensid-Lösungen im Frequenzbereich von 100 kHz bis 2 GHz*, Dissertation, Georg-August-Universität Göttingen (2003).
[77] R. M. Hill, 'Characterisation of Dielectric Loss in Solids and Liquids', Nature **275**, 96 (1978).
[78] R. M. Hill, 'Evaluation of Susceptibility Functions', Phys. Stat. Sol. (b) **103**, 319 (1981).
[79] K. Menzel, A. Rupprecht, and U. Kaatze, 'Hill-Type Ultrasonic Relaxation Spectra of Liquids', J. Acoust. Soc. Am. **104**, 2741 (1998).
[80] K. Giese, 'On the Numerical Evaluation of the Dielectric Relaxation Time Distribution Function from Permittivity Data', Adv. Molec. Relax. Processes **5**, 363 (1973).
[81] T. Telgmann and U. Kaatze, 'Monomer Exchange and Concentration Fluctuations of Micelles. Broad-Band Ultrasonic Spectrometry of the System Triethylene Glycol Monohexyl Ether/Water', J. Phys. Chem. A **104**, 1085 (2000).
[82] L. P. Kadanoff and J. Swift, 'Transport Coeffcients Near the Liquid-Gas Critical Point', Phys. Rev. **166**, 89 (1968).
[83] B. I. Halperin and P. C. Hohenberg, 'Scaling Laws for Dynamic Critical Phenomena', Phys. Rev. **177**, 952 (1969).
[84] R. A. Ferrell, 'Decoupled-Mode Dynamic Scaling Theory of the Binary-Liquid Phase Transition', Phys. Rev. Lett. **24**, 1169 (1970).
[85] P. C. Hohenberg and B. I. Halperin, 'Theory of Dynamical Critical Phenomena', Rev. Mod. Phys. **49**, 435 (1977).
[86] M. Brai and U. Kaatze, 'Ultrasonic and Hypersonic Relaxations of Monohydric Alcohol/Water Mixtures', J. Phys. Chem. **96**, 8946 (1992).
[87] B. Kühnel and U. Kaatze, 'Uncommon Ultrasonic Absorption Spectra of Tetraalkylammonium Bromides in Aqueous Solution', J. Phys. Chem. **100**, 19747 (1996).
[88] K. Menzel, A. Rupprecht, and U. Kaatze, 'Broad-Band Ultrasonic Spectrometry of C_iE_j/Water Mixtures. Precritical Behavior', J. Phys. Chem. B. **101**, 1255 (1997).
[89] A. Rupprecht and U. Kaatze, 'Model of Noncritical Concentration Fluctuations in Binary Liquids. Verification by Ultrasonic Spectrometry of Aqueous Systems and Evidence of Hydrophobic Effects', J. Phys. Chem. A **103**, 6485 (1999).
[90] A. Rupprecht and U. Kaatze, 'Solution Properties of Urea and Its Derivatives in Water: Evidence from Ultrasonic Relaxation Spectra', J. Phys. Chem. A **106**, 8850

(2002).
[91] V. P. Romanov and V. A. Solov'ev, 'Sound Absorption in Solutions', Sov. Phys. Acoust. **11**, 68 (1965).
[92] V. P. Romanov and S. V. Ul'yanov, 'Bulk Viscosity in Relaxing Media', Phys. A **201**, 527 (1993).
[93] V. P. Romanov and V. A. Solov'ev, 'Relaxation of the Ion Atmospheres and Sound Absorption in Electrolytes', Sov. Phys. Acoust. **19**, 550 (1974).
[94] C. J. Montrose and T. A. Litovitz, 'Structural-Relaxation Dynamics in Liquids', J. Acoust. Soc. Am. **47**, 1250 (1970).
[95] H. Endo, 'Sound Absorption Mechanism of an Aqueous Solution in Nonelectrolyte', J. Chem. Phys. **92**, 1986 (1990).
[96] K. Kawasaki, 'Kinetic Equations and Time Correlation Functions of Critical Fluctuations', Ann. Phys. (N. Y.) **61**, 1 (1970).
[97] J. K. Bhattacharjee and R. A. Ferrell, 'Dynamic Scaling Theory for the Critical Ultrasonic Attenuation in a Binary Liquid', Phys. Rev. A **24**, 1643 (1981).
[98] R. A. Ferrell and J. K. Bhattacharjee, 'General Dynamic Scaling Theory of Critical Ultrasonic Attenuation and Dispersion', Phys. Lett. A **86**, 109 (1981).
[99] R. A. Ferrell and J. K. Bhattacharjee, 'Dynamic Scaling Theory of the Critical Attenuation and Dispersion of Sound in a Classical Fluid: The Binary Liquid', Phys. Rev. A **31**, 1788 (1985).
[100] J. K. Bhattacharjee and R. A. Ferrell, 'Universality in the Critical Dynamics of Fluids: Ultrasonic Attenuation', Physica A **250**, 83 (1998).
[101] R. Folk and G. Moser, 'Frequency-Dependent Shear Viscosity, Sound Velocity, and Sound Attenuation Near the Critical Point in Liquids. I. Theoretical Results', Phys. Rev. E **57**, 683 (1998).
[102] R. Folk and G. Moser, 'Critical Dynamics in Mixtures', Phys. Rev. E **58**, 6246 (1998).
[103] R. Folk and G. Moser, 'Nonasymptotic Transport Properties in Fluids and Mixtures Near a Critical Point', Intern. J. Thermophys. **19**, 1003 (1998).
[104] R. Folk and G. Moser, 'Critical Sound in Fluids and Mixtures', Condens. Matter. Phys. **2**, 243 (1999).
[105] A. Onuki, 'Dynamic Equations and Bulk Viscosity Near the Gas-Liquid Critical Point', Phys. Rev. E **55**, 403 (1997).
[106] A. Onuki, 'Bulk Viscosity Near the Critical Point', J. Phys. Soc. Jpn. **66**, 511 (1997).
[107] R. Behrends and U. Kaatze, 'Scaling Frequency in the Critical Sound Attenuation of Binary Liquids', Europhys. Lett. **65**, 221 (2004).
[108] R. Behrends, U. Kaatze, and M. Schach, 'Scaling Function of the Critical Binary Mixture Methanol-Cyclohexane', J. Chem. Phys. **119**, 7957 (2003).
[109] I. Iwanowski, R. Behrends, and U. Kaatze, 'Critical Fluctuations Near the Consolute Point of n-Pentanol-Nitroethane. An Ultrasonic Spectrometry, Dynamic Light Scattering, and Shear Viscosity Study', J. Chem. Phys. **120**, 9192 (2004).
[110] R. Behrends, I. Iwanowski, M. Kosmowska, A. Szala, and U. Kaatze, 'Sound Attenuation, Shear Viscosity, and Mutual Diffusivity Behavior in the Nitroethane-Cyclohexane Critical Mixture', J. Chem. Phys. **121**, 5929 (2004).
[111] I. Iwanowski, K. Leluk, M. Rudowski, and U. Kaatze, 'Critical Dynamics of the Binary System Nitroethane/3-Methylpentane: Relaxation Rate and Scaling Function', J. Phys. Chem. A **110** 4313 (2006).
[112] I. Iwanowski, A. Sattarow, R. Behrends, S. Z. Mirzaev, and U. Kaatze, 'Dynamic Scaling of the Critical Mixture Methanol-Hexane', J. Chem. Phys. **124**, 144505 (2006).
[113] S. Z. Mirzaev, I. Iwanowski, and U. Kaatze, 'Dymanic Scaling and Background Relaxation in the Ultrasonic Spectra of the Ethanol-Dodecane Critical Mixture', Chem.

Phys. Lett. **435**, 263 (2007).
[114] S. Z. Mirzaev, I. Iwanowski, and U. Kaatze, 'Dynamic Scaling of the Critical Mixture Perfluoromethylcyclohexane Carbon Tetrachloride', J. Phys. D: Appl. Phys. **40**, 3248 (2007).
[115] H. C. Burstyn, J. V. Sengers, J. K. Bhattacharjee, and R. A. Ferrell, 'Dynamic Scaling Function for Critical Fluctuations in Classical FLuids', Phys. Rev. A **28**, 1567 (1983).
[116] J. K. Bhattacharjee, R. A. Ferrell, R. S. Basu, and J. V. Sengers, 'Crossover Function for the Critical Viscosity of a Classical Fluid', Phys. Rev. A **24**, 1469 (1981).
[117] J. C. Nienwouldt and J. V. Sengers, 'A Reevaluation of the Viscosity Exponent for Binary Mixtures Near the Consolute Point', J. Chem. Phys. **90**, 457 (1989).
[118] J. Luettmer-Strathmann, J. V. Sengers, and G. A. Olchowy, 'Non-Asymptotic Critical Behavior of the Transport Properties of Fluids', J. Chem. Phys. **103**, 7482 (1995).
[119] K. Kawasaki, 'Sound Attenuation and Dispersion Near the Gas-Liquid Point', Phys. Rev. A. **1**, 1750 (1970).
[120] A. Stein, J. C. Allegra, and G. F. Allen, 'Shear Viscosity of 3-Methylpentane-Nitroethane Near the Critical Point', J. Chem. Phys. **55**, 4265 (1971).
[121] C. W. Garland and G. Sanchez, 'Ultrasonic Study of Critical Behavior in the Binary Liquid 3-Methylpentane+Nitroethane', J. Chem. Phys. **79**, 3090 (1983).
[122] U. Kaatze and S. Z. Mirzaev, 'Slowing Down in Chemical Reactions. The Isobutyric Acid/Water System in the Critical Region', J. Phys. Chem. A **104**, 5430 (2000).
[123] S. Z. Mirzaev and U. Kaatze, 'Dynamic Scaling in the Ultrasonic Attenuation Spectra of Critical Binary Mixtures', Chem. Phys. Lett. **328**, 277 (2000).
[124] R. Behrends, T. Telgmann, and U. Kaatze, 'The Binary System Triethylamine-Water Near its Critical Consolute Point: An Ultrasonic Spectrometry, Dynamic Light Scattering, and Shear Viscosity Study', J. Chem. Phys. **117**, 9828 (2002).
[125] S. Z. Mirzaev, R. Behrends, T. Heimburg, J. Haller, and U. Kaatze, 'Critical Behavior of 2,6-Dimethylpyridine-Water: Measurements of Specific Heat, Dynamic Light Scattering, and Shear Viscosity', J. Chem. Phys. **124**, 144517 (2006).
[126] S. Z. Mirzaev, I. Iwanowski, M. Zaitdinov, and U. Kaatze, 'Critical Dynamics and Kinetics of Elementary Reactions of 2,6-Dimethylpyridine-Water', Chem. Phys. Lett. **431**, 308 (2006).
[127] I. Iwanowski and U. Kaatze, 'Dynamic Scaling and Slowing Down in Chemical Reactions of the Critical Triethylamine-Water System', J. Phys. Chem. B **111**, 1438 (2007).
[128] T. Telgmann and U. Kaatze, 'Monomer Exchange and Concentration Fluctuations in Poly(ethylene glycol) Monoalkyl Ether/Water Mixtures. Dependence upon Nonionic Surfactant Composition', J. Phys. Chem. A **104**, 4846 (2000).
[129] K. Menzel, S. Z. Mirzaev, and U. Kaatze, 'Crossover Behavior in Micellar Solutions with Critical Demixing Point: Broadband Ultrasonic Spectrometry of the Isobutoxyethanol-Water System', Phys. Rev. E **68**, 011501 (2003).
[130] U. Kaatze, 'Elementary Kinetics and Microdynamics of Aqueous Surfactants Systems', Uzbek. J. Phys. **5**, 75 (2003).
[131] I. Iwanowski, S. Z. Mirzaev, and U. Kaatze, 'Relaxation Rate in the Critical Dynamics of the Micellar Isobutoxyethanol-Water Sytem with Lower Consolute Point', Phys. Rev. E **73**, 061508 (2006).
[132] I. Procaccia and M. Gittermann, 'Slowing Down Chemical Reactions Near Thermodynamic Critical Points', Phys. Rev. Lett. **46**, 1163 (1981).
[133] H. G. E. Hentschel and I. Procaccia, 'Sound Attenuation in Critically Slowed-Down Chemically Reactive Mixtures', J. Chem. Phys. **76**, 666 (1982).

Oscillations, Waves and Interactions, pp. 405–434
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-15-7

Complex dynamics of nonlinear systems

Ulrich Parlitz

Drittes Physikalisches Institut, Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. Different nonlinear systems are presented exhibiting various dynamical phenomena, including bifurcations, chaotic dynamics, and synchronisation. Furthermore, methods for analysing, modelling and controlling complex dynamics are discussed. All these topics are illustrated with examples from work on nonlinear systems conducted at the DPI[1].

1 Introduction

Nonlinearity introduces a wealth of genuine dynamical phenomena, including multistability, different kinds of bifurcations, chaos and synchronisation, to mention some of the most important ones. Investigations on chaotic dynamics have shed new light on the notions of predictability and determinism, because the deterministic dynamics of even simple systems may be notoriously difficult to predict on short time scales and even impossible to forecast in the long run. The reason for this behaviour is *sensitive dependence on initial conditions*, where arbitrarily small perturbations of the system grow exponentially and thus become macroscopically relevant after some finite time. However, in spite of this extreme sensitivity chaotic systems can synchronise their aperiodic motion and, on the other hand, there exist different ways to suppress chaos by appropriate control methods.

Many aspects of nonlinear chaotic dynamics have been studied at the DPI[1] during the past decades. This research was initiated by Werner Lauterborn in the 1970s and 1980s when he published his seminal work on nonlinear bubble oscillations and chaos [1–3]. Later, many other nonlinear oscillators have been investigated in detail, and as a representative example we shall present some typical dynamical features of the driven Duffing oscillator in Sect. 2.1.

Another class of physical systems with very interesting nonlinear behaviour are lasers. Semiconductor lasers, for example, exhibit very complex chaotic dynamics when their emitted light is partly fed back by an external reflection. This phenomenon will be illustrated in Sect. 2.2 and in Sect. 5 synchronisation of two optically coupled chaotic semicondutor lasers is shown. Another type of lasers showing chaotic dynamics above some pump power threshold are frequency-doubled solid-state lasers. Since for many technical applications irregular intensity fluctuations are unwanted

[1]DPI = Drittes Physikalisches Institut = Third Physical Institute

here chaos control methods are of interest. As an example, in Sect. 6.1 multiple-delay feedback control is applied to a frequency-doubled Nd:YAG laser to suppress its chaotic fluctuations. This control method was developed at the DPI and has some promising features for stabilising and manipulating chaotic systems including complex spatio-temporal dynamics (Sect. 6.2). In Sect. 2.3 a combination of dynamical system governed by ordinary differential equations and an automaton switching between a finite number of states is presented. Such hybrid systems often occur in a technical context and may also exhibit chaotic dynamics. In Sect. 3 we briefly revisit the most important concepts for characterising chaotic dynamics: fractal dimensions and Lyapunov exponents. To investigate experimental systems specific methods for time series analysis are required. Following the long tradition of signal processing at the DPI, new methods for data analysis have been devised, implemented and applied. This topic is addressed in Sect. 4 where state space reconstruction (Sect. 4.1) and forecasting methods (Sect. 4.2) are discussed.

2 Nonlinear systems exhibiting complex dynamics

In this section three types of dynamical systems are presented which possess different dynamical features.

2.1 Nonlinear oscillators

A cornerstone of physics and engineering is the damped harmonic oscillator

$$\ddot{x} + d\dot{x} + \omega_0^2 x = f\cos(\omega t) \tag{1}$$

driven by some external periodic force $f\cos(\omega t)$.[2] It is well known that the solution of this ordinary differential equation (ODE)

$$x(t) = x_{hom}(t) + x_{inhom}(t) \tag{2}$$

consists of a general solution $x_{hom}(t)$ of the homogeneous equation ($f = 0$) converging to zero due to the damping and a special solution of the inhomogeneous system (1)

$$x_{inhom}(t) = a\cos(\omega t - \varphi) \tag{3}$$

with amplitude

$$a = \frac{f}{\sqrt{(\omega_0^2 - \omega^2)^2 + d^2\omega^2}} \tag{4}$$

and phase

$$\varphi = \arctan\left(\frac{d\omega}{\omega_0^2 - \omega^2}\right). \tag{5}$$

The amplitude a of the (asymptotic) oscillation of the driven linear oscillator is proportional to the driving force f and the shape and location of the resonance curve $a(\omega)$ which depends on the damping d and the eigenfrequency ω_0 as shown in Fig. 1.

[2] By rescaling time the eigenfrequency ω_0 can be eliminated (i. e., $\omega_0 \to 1$) but we shall keep it here for a more transparent interpretation of the results.

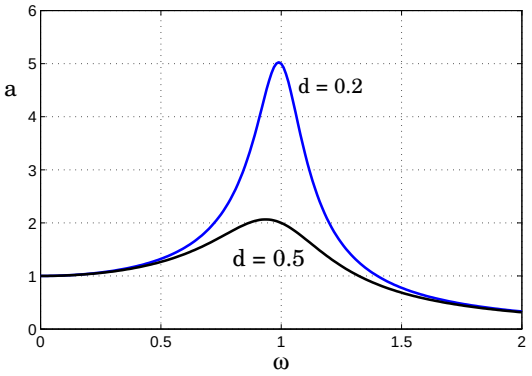


Figure 1. Resonance curves of the damped and driven harmonic oscillator Eq. (1) for high ($d = 0.5$) and low ($d = 0.2$) damping with $f = 1$ and $\omega_0 = 1$. The maximal amplitude a_{max} occurs at the resonance frequency $\omega_r = \sqrt{\omega_0^2 - d^2/2}$.

These relations completely describe the response of a (damped) harmonic oscillator to a sinusoidal excitation and can be extended to the case of a general periodic force using the superposition principle. Such a comprehensive treatment is possible only, because this system is linear. Many natural oscillators, however, are nonlinear and the question arises: how do nonlinearities change the dynamical behaviour? With the harmonic oscillator nonlinearity can enter in terms of a nonlinear restoring force and/or due to a nonlinear damping mechanism where the latter may render the oscillator a self-sustained system oscillating without any external driving.

A prototypical example of an oscillator with a nonlinear restoring force is the *Duffing oscillator*

$$\ddot{x} + d\dot{x} + \omega_0^2 x + \alpha x^3 = f\cos(\omega t)\,, \tag{6}$$

where the restoring force $\omega_0^2 x + \alpha x^3$ may be interpreted as a nonlinear approximation of a more general nonlinear force (Taylor expansion). By rescaling time t and x both coefficients ω_0^2 and α can be set to one and the resulting normalised Duffing equation reads

$$\ddot{x} + d\dot{x} + x + x^3 = f\cos(\omega t)\,. \tag{7}$$

The class of nonlinear oscillators with cubic restoring force is named after *Georg Duffing* engineer in Berlin in the beginning of the 20th century and published in 1918 a detailed study on *'Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung'*[3] [4].

As pointed out in the title of Duffing's book the "eigenfrequency" of nonlinear oscillators depends in general very much on the amplitude of the oscillation and actually provides a useful notion for relatively small amplitudes only, as we shall show in the following. This phenomenon is illustrated in Fig. 2 where amplitude resonance curves of the Duffing oscillator Eq. (7) are shown for $d = 0.2$ and three different driving amplitudes. For weak forcing with $f = 0.15$ the resulting (black) resonance curve still resembles the linear resonance curve (Fig. 1) but it is shifted towards higher frequencies and it bends to the right. If the driving amplitude is increased to $f = 0.3$ the resonance moves further to the right and the curve overhangs

[3] In English: Driven oscillations with variable eigenfrequencies and their technical importance.

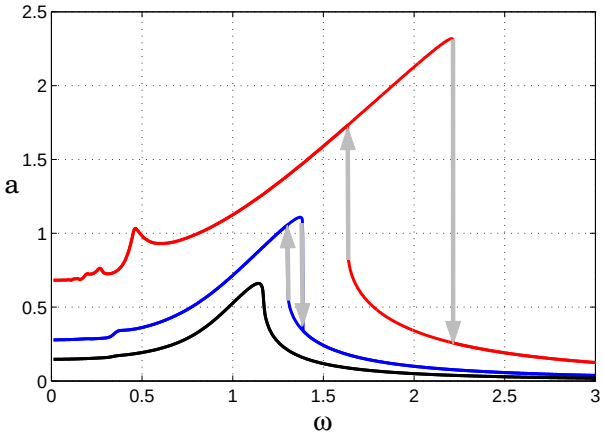


Figure 2. Amplitude resonance curve of the Duffing oscillator Eq. (7). Plotted is $a = \max(x(t))$ vs. ω for $d = 0.2$ and $f = 0.15$ (black curve), $f = 0.3$ (blue curve), and $f = 1$ (red curve). At some critical frequencies (bifurcation points) the oscillations loose their stability and the system undergoes a transient to another stable periodic solution as indicated by the arrows.

resulting in two branches. When slowly increasing the driving frequency ω starting from small values the oscillation amplitude grows until the end of the upper branch is reached. At that point the periodic oscillation looses its stability due to a saddle-node bifurcation and the driven oscillator converges to a stable periodic oscillation with much smaller amplitude corresponding to the lower branch. This transient is indicated by the grey arrow pointing downward and it typically takes several periods of the driving signal (i. e., it is *not* abrupt as may be suggested by the arrow). In the opposite direction, for decreasing excitation frequencies ω the system first follows the lower branch until this periodic oscillation becomes unstable and a transient to the upper branch occurs. In the frequency interval between both bifurcation points two stable periodic solutions exist for the oscillator and it depends on initial conditions whether the system exhibits the small or the large amplitude oscillations. Each stable periodic oscillation is associated with an attracting closed curve in state space called an *attractor* and we observe here a parameter interval with coexisting attractors. If the driving amplitude is increased to $f = 1$ this interval becomes larger and it is shifted further towards high driving frequencies (red curve in Fig. 2). Additionally, small peaks at low driving frequencies occur corresponding to nonlinear resonances. They bend to the left and they also overhang if the driving amplitude is sufficiently large. However, what is shown in Fig. 2 is just the tip of the iceberg and many additional coexisting attractors undergoing different types of bifurcations occur if the oscillator is forced into its full nonlinear regime. Before some of these features of the Duffing oscillator will briefly be discussed we want to have a look back again at Georg Duffing's pioneering work. He wrote that he first hoped to solve Eq. (7) using elliptic functions but then he realized soon that this is not possible. Then he applied perturbation theory and tested the resulting approximate solutions with mechanical experiments (also briefly presented in his book). His main interest was the shift of the resonance frequency and the occurrence of coexisting stable solutions including the resulting hysteresis phenomena (as shown in Fig. 2). Duffing's motivation for this study was mainly due to his interest in technical systems. In the introduction of his book [4] he describes some observations with synchronous electrical generators each

driven by a gas turbine. If the driving turbines did not run smoothly due to extra sparking the generators lost their synchrony and did not return to their previous oscillations when the extra sparking of the turbines was over. [4]

If the driving amplitude is further increased the cubic nonlinearity becomes more dominant and very different types of motion occur. Figure 3 shows five examples as time series, phase space projections and power spectra. The first time series in Fig. 3(a) is periodic with the same period as the driving signal. Therefore, this stable oscillation is also called a period-1 attractor. Figure 3(b) shows the corresponding trajectory in the $(x, \dot{x})$–plane (i. e., a projection of the attractor) where a red marker is plotted whenever a period $T = \frac{2\pi}{\omega}$ of the drive elapsed. This results in a stroboscopic phase portrait which is a convenient way for plotting Poincaré sections of periodically driven systems. Figure 3(c) shows a Fourier spectrum of the time series where only odd harmonics (multiples) of the fundamental frequency occur that coincides here with the driving frequency $\nu_0 = \omega/2\pi$. This feature is closely connected to the symmetry of the orbits in Fig. 3(b) that can be broken by a symmetry-breaking bifurcation as shown in Figs. 3(d–f). Symmetry breaking is a precursor of period-doubling because only asymmetric orbits can undergo a period-doubling bifurcation. An example for such a period-2 attractor is shown in Figs. 3(g–i) where now subharmonics occur in the spectrum (Fig. 3(i)) and two markers appear on the orbit (Fig. 3(h)). When the driving amplitude is increased furthermore a full period-doubling cascade takes place leading to chaotic dynamics as shown in Figs. 3(j–l). The oscillation is aperiodic (Fig. 3(j)) with a broadband spectrum (Fig. 3(l)) and a stroboscopic phase portrait (Fig. 3(k)) that constitutes a fractal set. Finally, Figs. 3(m–o) show a period-3 oscillation occuring for $f = 56$ which is an example for general period-m attractors that can be found for any m in some specific parts of the parameter space of Duffing's equation. Again, this period-3 attractor is symmetric (with odd harmonics of the fundamental frequency in the spectrum) and will undergo a symmetry-breaking bifurcation before entering a period-doubling cascade to chaos.

Figure 4 shows the Poincaré section of the chaotic attractor from Fig. 3(k) in more detail which possesses a self-similar structure as can be seen in the enlargement Fig. 4(b).

Poincaré cross sections are also an elegant way to visualise the parameter dependence of the dynamics. For this purpose a projection of the points in the Poincaré section (i. e., $z_1(n) = x(nT)$) is plotted vs. a control parameter that is varied in small steps. As initial conditions for the solutions of the equations of motions at a new parameter value the last computed state from the previous parameter is used. In this way, transients are kept short and one follows an attractor (and its metamorphoses

[4] In Ref. [4] on page 1 and 2 Duffing wrote: "Jene synchronen Drehstrommaschinen waren durch Gasmaschinen angetrieben. Die Frequenz der Antriebsimpulse und die sogenannte Eigenfrequenz waren genügend auseinander, so daß nur mäßige Pendelungen auftraten, wenn die Antriebsmaschinen im Beharrungszustande waren. Wurde dieser Beharrungszustand jedoch nur durch einige wenige heftigere Zündungen gestört, so wurden, auch nachdem die Verbrennungen wieder regelmäßig geworden war, die Pendelung immer noch größer und größer, so daß die Maschinen schließlich außer Tritt kamen. Nach den Ergebnissen der Theorie hätten, nach Eintreten der regelmäßigen Verbrennung, infolge der Dämpfung die Schwingungen im Laufe der Zeit wieder ihre normale Größe erhalten müssen."

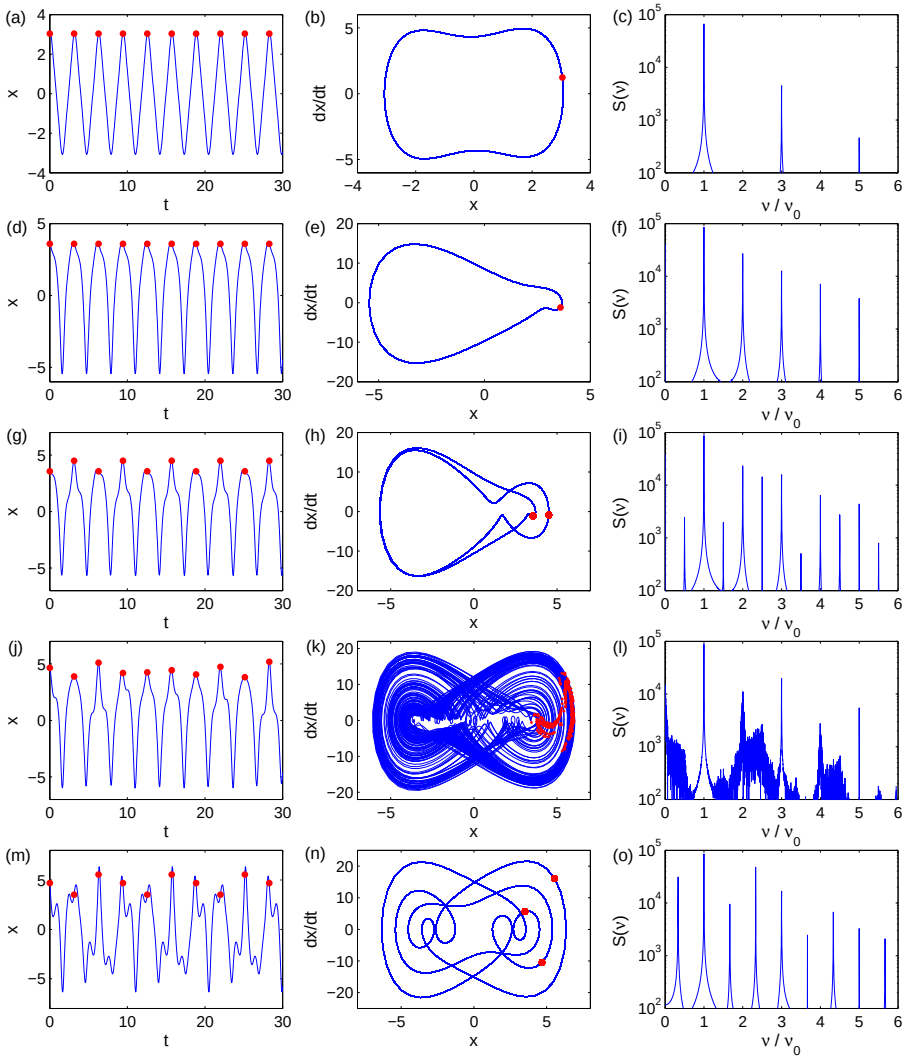


Figure 3. Time series, (stroboscopic) phase portraits, and power spectra of the Duffing oscillator Eq. (7) for $d = 0.2$, $\omega = 2$ and (a-c) $f = 10$ (symmetric period-1), (d-f) $f = 42$ (asymmetric period-1), (g-i) $f = 48$ (period-2), (j-l) $f = 54$ (chaos), and (m-o) $f = 56$ (symmetric period-3). Red markers are plotted whenever a period $T = 2\pi/\omega = 1/\nu_0$ of the drive elapsed.

and bifurcations) in parameter space. Figure 5(a) shows such a bifurcation diagram where the driving frequency ω is varied and in Fig. 5(b) the driving force f is used as a control parameter.

Although providing already a very detailed view of the bifurcation structure, bifurcation diagrams are limited to a single parameter axis. To get an even better overview

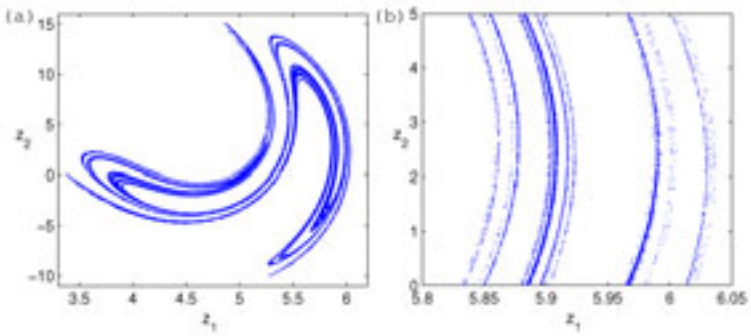


Figure 4. Poincaré section of a chaotic attractor of Duffing's equation (7) ($d = 0.2$, $\omega = 2$, $f = 54$) generated by stroboscopic sampling of $z_1(n) = x(nT)$ and $z_2(n) = \dot{x}(nT)$ with the driving period $T = 2\pi/\omega$. (a) Poincaré section of the full attractor (see Fig. 3k) and (b) section (zoom in) showing the underlying self-similar fractal structure.

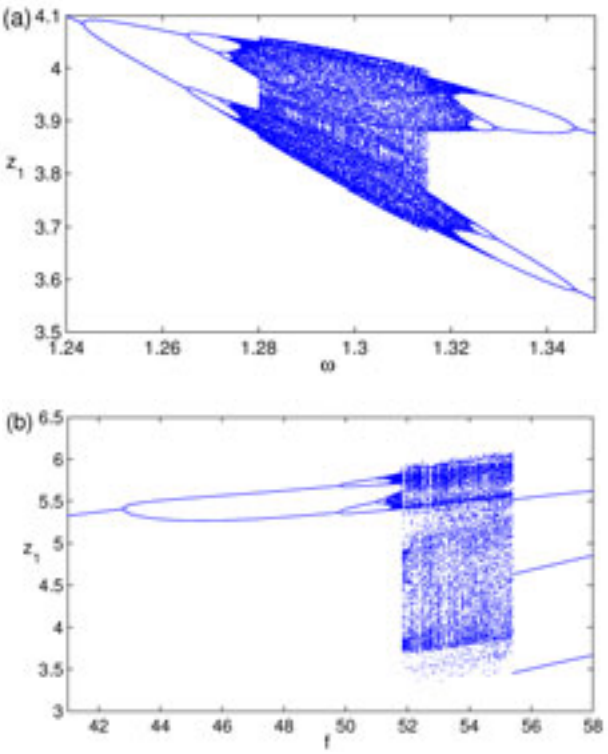


Figure 5. Bifurcation diagrams of the Duffing oscillator (7) with control parameter (a) driving frequency ω ($d = 0.2$, $f = 23$) and (b) driving amplitude f ($d = 0.2$, $\omega = 2$).

two-parameter studies are often desired and there are many bifurcation phenomena requiring (at least) two control parameters (called codimension-two bifurcations). Two-dimensional charts of the parameter space showing parameter combinations where bifurcations occur are called *phase diagrams*. In these diagrams (codimension-one) bifurcations are associated with curves as can be seen in Fig. 6 for the Duffing oscillator. The damping constant $d = 0.2$ is kept fixed and the driving frequency ω and the driving amplitude f are plotted along logarithmic axes, to emphasise the repeated structure of bifurcation curves. This superstructure of the bifurcation set [5] is closely connected to the nonlinear resonances of the oscillator (partly visible in Fig. 2) and the torsion of the flow along the periodic orbits [6]. In the parameter regions coloured in orange asymmetric period-1 oscillations exist (similar to the example shown in Fig. 3(d–f)) and in the yellow regions period-doubling cascades and chaos occur. There is some very characteristic pattern of bifurcation curves (not shown here) that occurs not only in all resonances of Duffing's equation but also with many other driven nonlinear oscillators [7,8]. Furthermore, the phase diagram in Fig. 3 doesn't show all bifurcations in the chosen section of the parameter space. There are, for example, coexisting attractors in phase space which independently undergo their own bifurcation scenarios when ω and f are varied. A complete study and smart ways for visualising the plethora of attractors and bifurcations of Duffing's equation (and other nonlinear oscillators) remains a challenge.

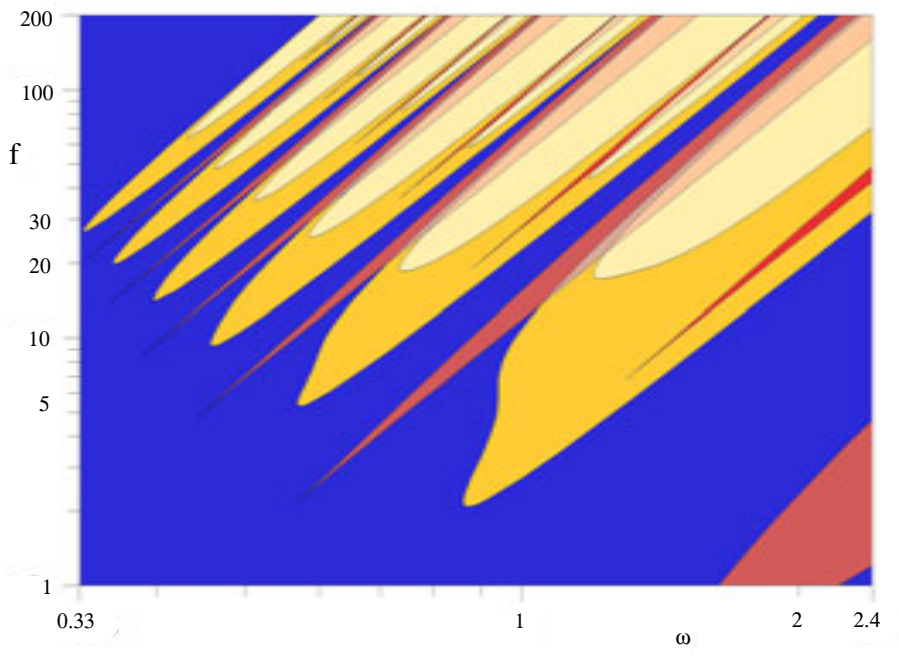


Figure 6. Phase diagram of the Duffing oscillator Eq. (7) for $d = 0.2$ [9].

2.2 Semiconductor lasers with external cavities

Lasers are known as prototypical system with extremely well-ordered dynamics. This order, however, can be destroyed by nonlinearities and feedback. A typical example for this phenomenon are semiconductor lasers with external cavities [10]. Pointing the laser at a mirror, such that a fraction of its own light output is led back into its internal resonator, the laser's intensity will start to fluctuate and, depending on a number of parameters, different interesting dynamics can occur, including high-dimensional chaos.

These phenomena are known since 1977, when Risch and Voumand [11] were the first to describe the so called *low-frequency fluctuations* (LFFs) of a semiconductor laser. They occur for very small reflectivities and pump currents only slightly above the threshold current. The light output of LFFs is characterised by frequent and very sudden *power drop-outs*, each followed by a relatively slow recreation of the light intensity. Modulated onto them is a fast oscillation that can usually only be seen low-pass filtered due to the finite response time of the photo diode used to capture the intensity and limited transfer functions of subsequent amplifiers and oscilloscopes. The frequency of the power drop-outs is only 3–30 MHz, which is relatively slow compared to the fast oscillations that are in the range of several GHz.

The first model describing the dynamics of semiconductor lasers with optical feedback was proposed by Lang and Kobayashi [12] in 1980. They extended the well-known semiconductor rate equations by a feedback term, which leads to a system of equations known as the *Lang-Kobayashi equations* (LKEs):

$$\begin{aligned} \dot{E}_0(t) &= \frac{1}{2} G_N n(t) E_0(t) + \kappa E_0(t-\tau) \cos[\omega_0 \tau + \phi(t) - \phi(t-\tau)] \\ \dot{\phi}(t) &= \frac{1}{2} \alpha G_N n(t) - \kappa \frac{E_0(t-\tau)}{E_0(t)} \sin[\omega_0 \tau + \phi(t) - \phi(t-\tau)] \\ \dot{n}(t) &= (p-1) J_{th} - \gamma n(t) - [\Gamma + G_N n(t)] E_0^2(t)\,, \end{aligned} \tag{8}$$

where κ is the differential feedback rate. The equation for the complex electric field $E(t) = E_0(t) \exp(\mathrm{i}[\omega_0 t + \phi(t)])$ has been split into two one-dimensional real equations, and $n(t) = N(t) - N_{sol}$ is the carrier number above the value N_{sol} of the unperturbed, "solitary" semiconductor laser with no optical feedback. τ is the light round-trip time in the external resonator. All other parameters are explained in Table 1. In general, the LKEs also contain a stochastic term for describing spontaneous emission that has been omitted here. Since Eq. (8) is a delay differential equation it describes a dynamical system with an infinite-dimensional state space where very complex high-dimensional chaotic dynamics may occur (and does occur!).

Using these equations, Lang and Kobayashi were able to simulate the phenomena they discovered for small reflectivities and laser-mirror distances of 1–2 cm, such as multistability or hysteresis.

The origin of the LFF dynamics remained unclear for many years. Fujiwara et al. [13] suggested the LFFs to result from a decreased relaxation oscillation frequency. Henry and Kazarinov [14] assumed a stable resonator mode out of which the laser is randomly kicked by spontaneous emission noise, causing power drop-outs, and Hohl

G_N	$2.142 \cdot 10^{-5}\,\mathrm{ns}^{-1}$	differential optical gain
α	5.0	linewidth enhancement factor
γ	$0.909\,\mathrm{ns}^{-1}$	carrier loss rate
Γ	$0.357\,\mathrm{ps}^{-1}$	photon loss rate
J_{th}	$1.552 \cdot 10^{8}\,\mathrm{ns}^{-1}$	threshold current density
p	1.02	pump current density over J_{th}
$2\pi c/\omega_0$	635 nm	solitary laser wavelength
N_{sol}	$1.707 \cdot 10^{8}$	solitary laser carrier number
κ	$10^{11}\,\mathrm{s}^{-1}$	feedback rate
τ	10 ns	external cavitiy round trip time

Table 1. Typical parameter values of the Lang-Kobayashi equations (8).

et al. [15] showed that characteristics and statistics of the LFFs is indeed influenced by this noise. Mørk et al. [16] assumed the laser to become bistable due to the feedback, such that the spontanous emission noise would cause a mode hopping between these two states. In 1994, Sano [17] showed that the LFF dynamics could be simulated by the deterministic LKEs (8). Those simulations also revealed the frequency of the fast oscillations mentioned above that were eventually visualised experimentally by Fischer et al. [18] in 1996. In 1998 Ahlers et al. [19] showed at the DPI that chaotic LFF dynamics generated by the LKEs may possess many positive Lyapunov exponents (hyperchaos) and can be synchronised by optical coupling (see Sect. 5.3). On the other hand, the chaotic fluctuations disappear if additional external cavities are used with suitably chosen lengths and resulting delay times.

LFFs were generally assumed to be a phenomenon of only low pump currents, until in 1997 Pan et al. [20] were the first to show both experimentally and numerically that in case of larger feedback rates they occur for currents well above the laser threshold as well. In this regime, power drop-outs turn into *power jump-ups*. They appear like inverted power drop-outs, but their modelling is more complicated [21]. and requires multiple reflections to be taken into account [22].

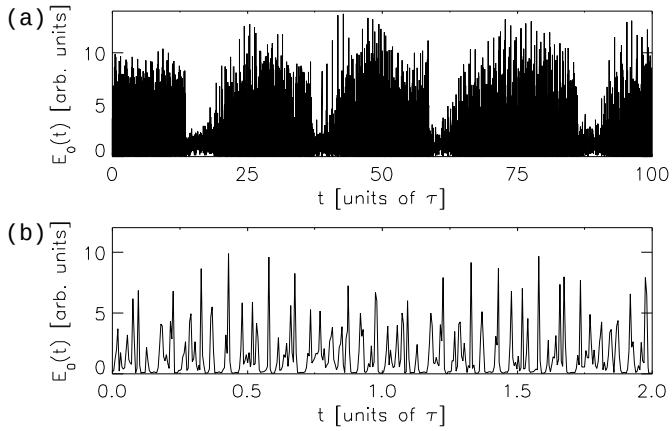


Figure 7. Low frequency fluctuations of a semiconductor laser with weak external optical feedback. Simulation with the Lang-Kobayashi equations (8) and the parameters given in Table 1 [19].

2.3 Hybrid systems

The third example of a dynamical system exhibiting chaotic dynamics is a hybrid system consisting of a set of ordinary differential equations (ODEs) coupled to an automaton (i. e., a finite state machine). This kind of systems occurs, for example, whenever the dynamics is governed by some switching rules. With Karsten Peters we studied simple hybrid systems in order to understand irregular behaviour in production systems as used in large factories [23]. Of course, such systems are perturbed by many stochastic influences (e. g., human behaviour, accidents, supply problems, ...) but one may ask the question whether irregular behaviour in a production line may also originate from its internal rules and operating conditions. In other words, does an ideal production process without any stochastic influences always operate properly with some constant or periodic output? To address this question we investigated a simple hybrid system representing some unit in a larger production line. As illustrated in Fig. 8 this unit consists of a server S processing three types of workpieces coming from previous production units P_1, P_2, and P_3. Since at a given instant of time the server can process the input from one of the production units, only, some switching rules are required and the outputs of the P_i have to be stored in some buffers. Let x_1, x_2, and x_3 denote the contents of the buffers belonging to P_1, P_2, and P_3, respectively. If each of the production units P_i provides constant output with some rate f_i the buffer contents x_i will increase with $\dot{x}_i = f_i$. On the other hand, the buffers are emptied by the server which formally can be described by emptying rates e_i. The ODE-system describing the buffer contents is thus given by

$$\begin{aligned} \dot{x}_1 &= f_1 - e_1 \\ \dot{x}_2 &= f_2 - e_2 \\ \dot{x}_3 &= f_3 - e_3 \end{aligned} \tag{9}$$

where the filling rates f_i and the emptying rates e_i should fulfill a balancing condition $f_1 + f_2 + f_3 = c = e_1 + e_2 + e_3$ to avoid overflow or complete emptying of buffers. The total input rate c is constant if all rates f_i are constant and without loss of generality we shall set $c = 1$ in the following. As already mentioned at a given time the server can process only input from one of the production units P_i. Therefore, the emptying rates e_i are actually functions of time and only one of them is larger than zero corresponding to the P_i which is currently emptied by the server S. This is the point where some switching rules have to be defined that state which P_i has to be served ($e_i(t) > 0$) at a given time. A possible set of rules is the following:

- If a buffer content x_k reaches some maximal value b then switch server S to that buffer to empty it ($\rightarrow e_k = 1$).
- If the buffer that is currently processed by S is empty ($x_k(t) = 0$) then switch server S to the next buffer ($\rightarrow e_{k+1} = 1$) in a cyclic order.

Here we introduced the maximal buffer content b which is an important parameter from the practical as well as the dynamical point of view. Of course, any manufacturer will try to keep buffers small. But it turned out that the dynamics very strongly

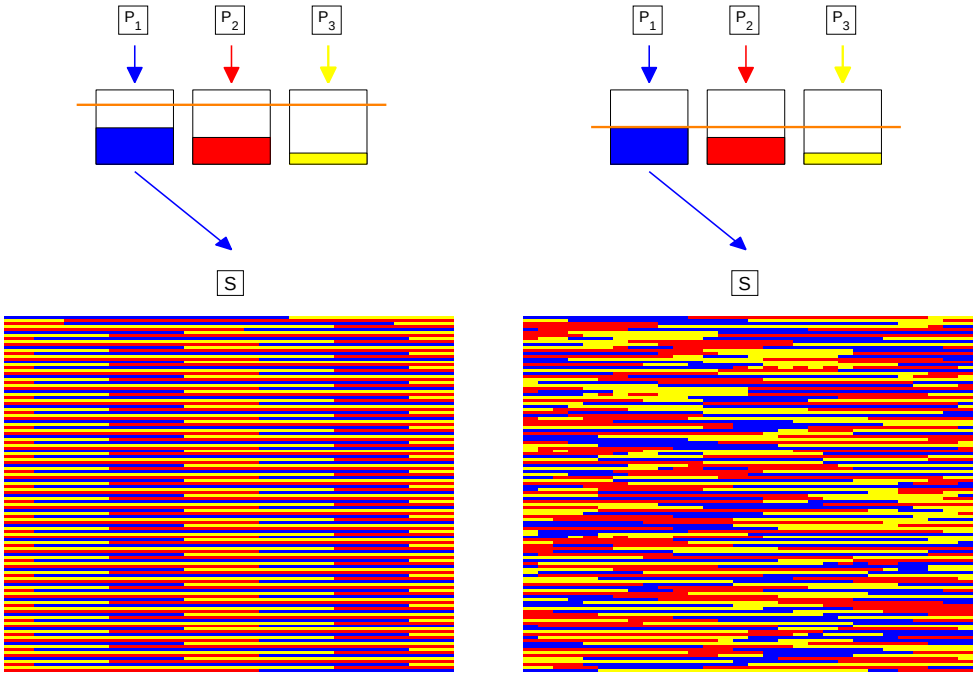


Figure 8. Tank system describing a logistic process where three production units P_1, P_2, and P_3 deliver some inputs (e. g., workpieces) that are stored in buffers and processed by a subsequent server S that switches between the buffers to empty them. The filling is governed by Eqs. (9) with $f_i = 1/3$ following the switching rule stated in the text. If the buffer size b (indicated by the horizontal orange line) is large the resulting dynamics is periodic (left column, $b = 0.8$). Too small buffers, however, lead to chaotic switching as illustrated with the example shown in the right column for $b = 0.5$.

depends on b. This is illustrated in Fig. 8. In the left column a case is shown where $b = 0.8$ is relatively large. The system started with a configuration shown in the top row where the server processes the blue buffer P_1. When P_1 is emptied S switches to the yellow buffer P_3 etc. following the rules stated above. The switching process is shown row by row in the lower left diagram starting in the upper left corner. As can be seen a periodic state is reached after some short transient visible in the first six rows. The right column of Fig. 8 shows what happens if the buffer size is reduced to $b = 0.5$. Now the server doesn't operate periodically anymore but switches chaotically between the buffers of the production units.

This shows that irregular dynamics can occur in (production) logistic processes even if they are not randomly perturbed. Here, the transition to chaos consists of interesting (non-standard) bifurcation scenarios and of course the type of dynamics has also some influence on the throughput of the whole unit [23,24]. An example for a hybrid (or tank) system occuring in physics is discussed in Ref. [25] where this formalism is used to describe front dynamics in semiconductors.

3 Characterising complex dynamics

The most important quantities for characterising chaotic dynamics are attractor dimensions and Lyapunov exponents that we shall briefly introduce in the following.

3.1 Fractal dimensions

The (fractal) dimension of an attractor quantifies its complexity and gives a lower bound for the number of equations or variables needed for modelling the underlying dynamical process.

The simplest concept of a fractal dimension is the *box-counting dimension* (or *capacity dimension*). There, the point set (attractor) to be characterised is covered with N d-dimensional hypercubes of size ε. The smaller ε is, the more cubes are necessary and the scaling

$$N(\varepsilon) \propto \left(\frac{1}{\varepsilon}\right)^{D_B} \tag{10}$$

of the number of cubes $N(\varepsilon)$ with the size ε provides, in the limit of infinitesimally boxes, the *box-counting dimension*

$$D_B = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)} \,. \tag{11}$$

With the box-counting dimension a hypercube is counted if it already contains just a single point of the set to be described. In general, however, also the local density of points is of interest, given by the probability p_i to find a point in cube number i. This probability can be estimated by the relative number of points falling in box i and depends on the size ε. For a general point set (attractor) covered with $N(\varepsilon)$ d-dimensional hypercubes of size ε we obtain in this way the *Rényi information of order q*

$$I_q = \frac{1}{1-q} \ln \sum_{i=1}^{N(\varepsilon)} p_i^q \,, \tag{12}$$

which is used to define the *generalised (Rényi) dimension of order q*

$$D_q = \lim_{\varepsilon \to 0} \frac{I_q}{\ln(1/\varepsilon)} \,. \tag{13}$$

Note that q can be any real number, i. e., Eq. (13) describes an infinite family of dimensions. For $q = 0$ the Rényi information I_0 equals $\ln N(\varepsilon)$ and Eq. (13) coincides with the definition of the box-counting dimension, $D_0 = D_B$.

From the infinite family of (generalised) dimensions D_q the *correlation dimension* D_C introduced by Grassberger and Procaccia [26] is often used for analysing strange (chaotic) attractors. The correlation dimension is given by the scaling

$$C(r) \propto r^{D_C} \tag{14}$$

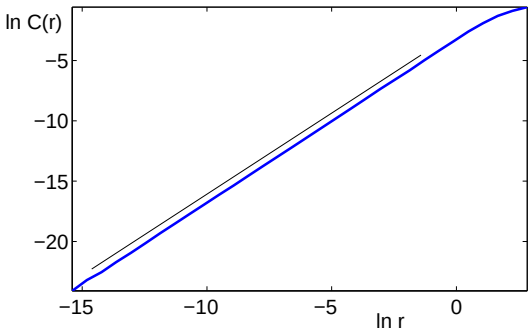


Figure 9. Correlation sum Eq. (15) vs. radius r for the Poincaré cross section Fig. 4 of the chaotic attractor of Duffing's equation computed using 100000 points. The slope of ≈ 1.34 represented by the (black) line segment is an approximation of the correlation dimension D_C, Eq. (16).

of the *correlation sum*

$$C(r) = \frac{2}{(N)(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{i-1} H(r - \|\mathbf{y}^i - \mathbf{y}^j\|) \tag{15}$$

that counts the relative number of neighbouring points $\mathbf{y}^i$ and $\mathbf{y}^j$ closer than r (H is the Heaviside function with $H(x) = 1$ for $x > 0$ and zero elsewhere). The correlation dimension

$$D_C = \lim_{r \to 0} \frac{\ln C(r)}{\ln r} \tag{16}$$

describes the scaling in the limit $r \to 0$ and equals the Rényi dimension D_2.

Since any numerical simulation or experimental measurement provides *finite* data sets, only, the limit $r \to 0$ cannot be carried out in numerical computations of D_C, but only the corresponding scaling behaviour

$$\ln C(r) \approx D_C \ln r \tag{17}$$

can be exploited to estimate D_C as a slope in a log-log-diagram. This approach is illustrated in Fig. 9 showing $\ln C(r)$ vs. $\ln r$ for the Poincaré section of the chaotic Duffing attractor in Fig. 4. The slope in a suitable intermediate range of r gives the fractal dimension of the Poincaré section of the chaotic Duffing attractor (here estimated as $D_C \approx 1.34$).

Each point in the Poincaré cross section corresponds to a one-dimensional trajectory (segment) starting at that point. Therefore, the dimension estimate in the Poincaré section has to be increased by one to obtain the (correlation) dimension of the full chaotic attractor (here $D_C \approx 2.34$).

The definition of the correlation dimension (16) is based on a given radius r and therefore a *fixed size* approach. As an alternative one may also use a *fixed mass* method to estimate the dimension of the attractor as was suggested by Badii and Politi [27,28] and Grassberger [29]. In this case the k nearest neighbours of each reference point $\mathbf{y}^n$ and the radius $r^n = r(k)$ of this cloud of k points are determined. For the limit $k/N \to 0$ one obtains, for example, an approximation of the box-counting dimension

$$D_B \approx -\frac{\log N}{\log \frac{1}{N} \sum_{n=1}^{N} r^n} \,. \tag{18}$$

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

All methods for computing fractal dimensions from (large) data sets can be considerably accelerated by using fast search algorithms [30] for the nearest neighbours of the data points. Algorithms for these tasks and for estimating many other useful characteristics of nonlinear systems were implemented at the DPI and are publicly available in the Matlab^{TM} toolbox TSTOOL[5].

More details about dimension estimation methods, their possible pitfalls, extensions, and further references are given in many review articles and textbooks [31–40].

3.2 Lyapunov exponents

Lyapunov 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. The full set of Lyapunov exponents of a d-dimensional system constitutes the *Lyapunov spectrum* which is an ordered set of real numbers $\{\lambda_1, \lambda_2, \ldots, \lambda_d\}$ with $\lambda_i \geq \lambda_{i+1}$.

When the largest Lyapunov exponent λ_1 is positive, the system is said to be chaotic and it shows sensitive dependence on initial conditions.

The meaning of the Lyapunov exponents is illustrated in Fig. 10. An infinitesimally small ball of initial conditions forming neighbouring points of some reference state is transformed into an ellipsoid due to the temporal evolution of the system (linearized around the trajectory of the reference state). The principal axes of this ellipsoid grow or shrink proportial to $\exp(\lambda_i t)$.

For an exact definition of the Lyapunov exponents and computational details see Geist et al. [41] or Abarbanel [38].

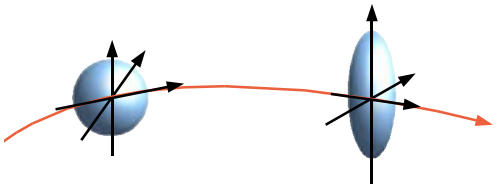


Figure 10. Illustration of the local temporal evolution of a ball of neighbouring states evolving into an ellipsoid with principle axes whose lengths are proportional to $\exp(\lambda_i t)$.

4 Time series analysis

In mathematical models of dynamical systems the dynamics is described in their state space, whose (integer) dimension is given by the number of the dependent variables of the model. In experiments, however, often just one variable is measured as a function of time, and the state space is usually not known. How, then, to arrive at the attractor that may characterise the system? This gap between the theoretical notions and observable quantities was filled in 1980 when Packard, Crutchfield, Farmer, and Shaw [42] published their fundamental paper "Geometry from a time

[5]TSTOOL URL: `//http:www.physik3.gwdg.de/tstool/`

series" where for the first time state space reconstruction methods were applied to scalar time series. A mathematical justification of this approach was given by Takens [43] at about the same time. He proved that it is possible to (re)construct, from a scalar time series only, a new state space that is diffeomorphically equivalent to the (in general unknown) original state space of the experimental system. Based on these reconstructed states the time series can then be analysed from the point view of (deterministic) nonlinear dynamics. It is possible to model and predict the underlying dynamics and to characterize the dynamics in terms of dimensions and Lyapunov exponents, for example. In the following we will address some of these issues. For further reading we refer to other review articles [35–39] and implementations of many time series algorithms can be found in the TSTOOL box already mentioned in Sect. 3.1.

4.1 State space reconstruction

Essentially two methods for reconstructing the state space from scalar time series are available: *delay coordinates* and *derivative coordinates*. Derivative coordinates were used by Packard et al. [42] 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.

Let M be a smooth (C^2) m–dimensional manifold in the state space in which the dynamics of interest takes place and let $\phi^t : M \to M$ be the corresponding flow describing the temporal evolution of states in M. 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*

$$\begin{array}{rccl} F: & M & \to & \mathbb{R}^d \\ & \mathbf{x} & \mapsto & \mathbf{y} = F(\mathbf{x}) = (s(t), s(t-t_l), ..., s(t-(d-1)t_l) \end{array}$$

that maps a state $\mathbf{x}$ from the original state space M to a point $\mathbf{y}$ in a *reconstructed state space* $\mathbb{R}^d$, where d is the *embedding dimension* and t_l gives the *delay time* (or *lag*) used. Figure 11 shows a visualisation of this construction. Takens [43] 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 generalised by Sauer, Yorke and Casdagli [44,45] who 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 guarantees only for very large embedding dimensions d (e. g., $d \geq 201$) the existence of a diffeomorphic equivalence, whereas with the condition of Sauer et al. a much smaller d will suffice (e. g., $d > 10$). 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

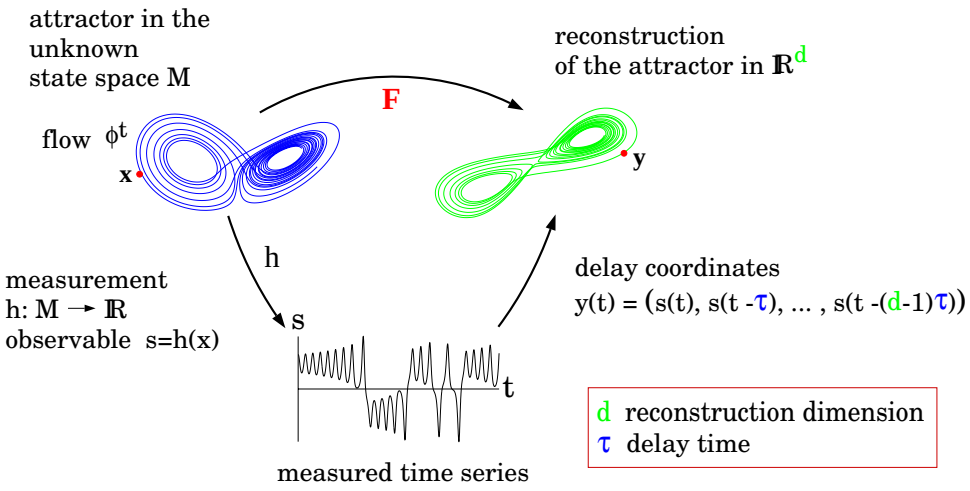


Figure 11. Delay reconstruction of states from scalar time series.

about the reconstruction of states, in particular in the presence of noise, may be found in Refs. [46,47].

If the data are measured with a high sampling rate *Broomhead-King-coordinates* [48, 49] may be advantageous. With this method a very high-dimensional reconstruction is used and then a new coordinate system is introduced 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 (states). This new coordinate system is based on a Karhunen-Loève transformation [6] 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 Ref. [50].

For time series that consist of a sequence of sharp spikes (e. g. from firing neurons) delay embedding may lead to very inhomogeneous sets of points in the reconstructed state space that are difficult to analyse. As an alternative one may use in this case the time intervals between the spikes as components for (re-)constructed state vectors [51–53].

4.2 Forecasting and Modelling

After successful state space reconstruction one can approximate the dynamics in reconstruction space to forecast or control the underlying dynamical process. Very simple but efficient algorithms for nonlinear prediction are nearest neighbours methods (also called local models). Let's assume that we want to forecast the future evolution of a given (reconstructed) state for some time horizon T. Available (i. e.,

[6] Also called *Proper Orthogonal Decomposition* (POD) or *Principle Component Analysis* (PCA).

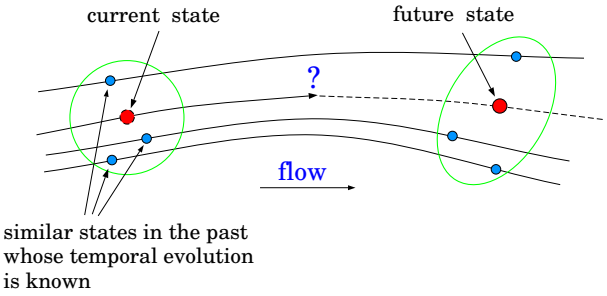


Figure 12. Local modelling using nearest neighbours in (reconstructed) state space.

known) are some neighbouring (reconstructed) states of this reference state that occurred (in the given time series) in the past such that their evolution over a period of time T is already known (as illustrated in Fig. 12). If the dynamical flow in (reconstructed) state space is continuous then the future values of the neighbouring states provide good approximations of the future evolution of the reference state. This is the main idea of the local approach and there are several options for further improving its performance [54]. Local modelling can also be applied to complex extendend system if a suitable state space reconstruction method is used [55].

An alternative to local modelling are global models, for example given as a superposition of nonlinear basis functions. Such models have been employed to describe not only the dynamics of a given system but also its parameter dependence [56]. If models with good generalisation capabilities are required (i. e., models with good performance on data not seen during the learning process) it is often advantageous to use not a single type of model but an ensemble of different models. Averaging their individual forecasts provides in most cases better results (on average) than any single model [56,57]. A Matlab^{TM} toolbox ENTOOL for such ensemble modelling was developed by former DPI students Christian Merkwirth and Jörg Wichard.[7]

5 Synchronisation of chaotic dynamics

Synchronisation of periodic signals is a well-known phenomenon in physics, engineering and many other scientific disciplines. It was first investigated in 1665 by Christiaan Huygens who observed that two pendulum clocks hanging at the same beam of his room oscillated in exact synchrony [58,59]. Huygens made experiments with his clocks and found that the synchronisation originates from invisible vibrations of the beam enabling some interaction between both oscillators. He reported his findings on the "sympathy of two clocks" (as he called it) at the Royal Society of London but it took more than 200 years before research on synchronisation was continued. J. W. Strutt (Lord Rayleigh) described in the middle of the 19th century that two (similar) organ pipes sound unisono if placed close together so that they can interact acoustically [60].

[7] It is available at `http://zti.if.uj.edu.pl/~merkwirth/entool.htm`.

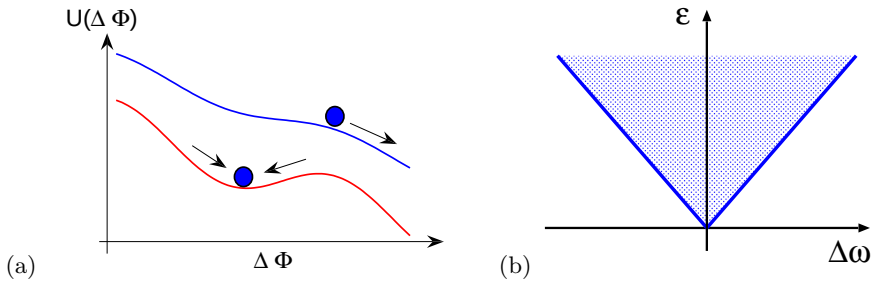


Figure 13. (a) Potential $U(\Delta\phi) = -\Delta\omega\Delta\phi - \varepsilon\cos(\Delta\phi)$ of the Adler Eq. (19) for weak ($|\varepsilon| < |\Delta\omega|$, blue curve) and strong coupling ($|\varepsilon| > |\Delta\omega|$, red curve), where $\Delta\phi(t)$ converges to local minima. (b) Stability region (Arnol'd tongue, shaded) where $\Delta\phi(t)$ converges to some fixed value and both oscillators synchronise.

5.1 Synchronisation of periodic oscillations

Modern research on synchronisation began in the 1920s and again, it were technical systems (vacuum tube oscillators) where synchronisation phenomena were observed and investigated in detail by E. V. Appleton [61] and B. van der Pol [62] (based on previous work and a patent of W. H. Eccles and J. H.Vincent) [58]. R. Adler [63] showed in 1945 for a general pair of weakly coupled periodic oscillators that their phase difference $\Delta\phi = \phi_1 - \phi_2$ is governed by a differential equation

$$\frac{d\Delta\phi}{dt} = \Delta\omega - \varepsilon\sin\Delta\varphi\,, \tag{19}$$

where $\Delta\omega = \omega_1 - \omega_2$ denotes the (small) frequency mismatch between the free-running oscillators (with individual frequencies ω_1 and ω_2) and ε is the (small) coupling strength. Stability analysis shows that $\Delta\phi$ grows unbounded if the coupling is weak ($|\varepsilon| < |\Delta\omega|$) but converges to a fixed value if the coupling exceeds some threshold ($|\varepsilon| > |\Delta\omega|$). This dynamical behaviour can also be visualised as motion of a particle in a potential $U(\Delta\phi) = -\Delta\omega\Delta\phi - \varepsilon\cos(\Delta\phi)$ and results in a wedge-shaped stability region (Arnol'd tongue) in the $(\Delta\omega, \varepsilon)$–parameter space where synchronisation (i. e., entrainment) occurs (see Fig. 13). The same synchronisation analysis holds for periodically driven systems.

5.2 Phase synchronisation of chaotic oscillations

In Adler's equation both oscillators are described by their phases, an approximation that is valid for weakly coupled periodic systems [58]. However, synchronisation phenomena are not restricted to this class of dynamical systems but occur also for coupled or driven chaotic oscillators [64–68].

As an example we investigated with Lutz Junge an analog circuit implementa-

tion [69] of two uni-directionally coupled Rössler systems (20) and (21)

$$\begin{aligned} \alpha\dot{x}_1 &= 2 + x_1(x_2 - 4) \\ \alpha\dot{x}_2 &= -x_1 - \omega_1 x_3 \\ \alpha\dot{x}_3 &= \omega_1 x_2 + 0.412 x_3 \end{aligned} \tag{20}$$

$$\begin{aligned} \alpha\dot{y}_1 &= 2 + y_1(y_2 - 4) \\ \alpha\dot{y}_2 &= -y_1 - \omega_2 y_3 \\ \alpha\dot{y}_3 &= \omega_2 y_2 + 0.412 y_3 + c(x_3 - y_3). \end{aligned} \tag{21}$$

Both Rössler systems exhibit chaotic oscillations when uncoupled ($c = 0$), but with different mean frequencies given by the parameters $\omega_1 = 1$ and $\omega_2 = 1.1$. The parameter $\alpha = 0.013$ is a (time) scaling factor due to the hardware implementation. In order to obtain a description in terms of phase variables, attractors have been reconstructed from time series (16 bit resolution, 1 kHz sampling frequency) of the x_2 and the y_2 variable using the method of delays (see Sect. 4.1) and are shown for $c = 0$ in Fig. 14. From these reconstructions phases (angles) $\phi_1(t)$ and $\phi_2(t)$ and mean rotation frequencies

$$\Omega_i = \lim_{t\to\infty} \frac{\phi_i(t)}{t} \tag{22}$$

were computed using polar coordinates centered in the 'hole' of each reconstructed attractor. If both Rössler systems are uncoupled their mean rotation frequencies Ω_1 and Ω_2 are different due to the different parameters $\omega_1 = 1$ and $\omega_2 = 1.1$ in Eqs. (20) and (21). This difference still exists for sufficiently small values of the coupling parameter c as can be seen in Fig. 15 where the mean rotation frequencies Ω_1 (dashed) and Ω_2 (solid) are plotted vs. c. At $c \approx 0.18$ the response system undergoes a transition to a new phase synchronised state where the mean rotation frequencies of the drive (20) and the response system (21) coincide.

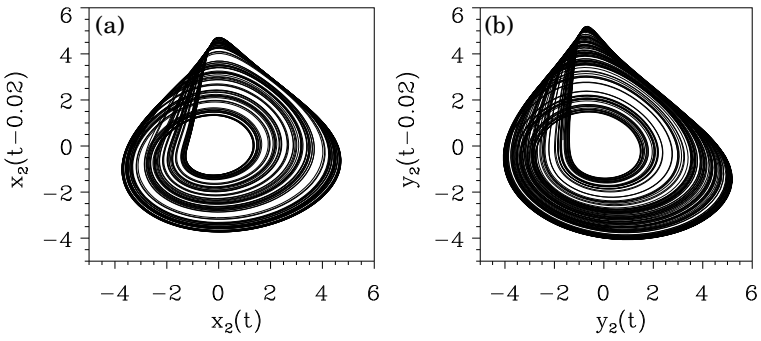


Figure 14. Delay reconstruction of the attractors of the drive (a) and the response system (b) given by Eqs.(20) and (21), respectively. Both time series were generated experimentally using an analog computer. The mean rotation frequencies are $\Omega_1 = 11.82$ Hz (a) and $\Omega_2 = 13.62$ Hz (b) [69].

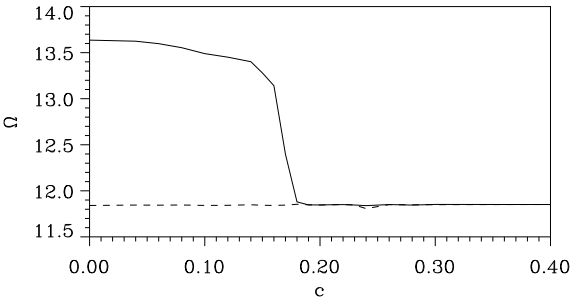


Figure 15. Mean rotation frequencies Ω_1 (dashed) and Ω_2 (solid) vs. the coupling parameter c. For $c > 0.18$ phase synchronisation occurs and both rotation frequencies coincide [69].

Figure 16 shows the phase difference $\Delta\phi(t) = \phi_1(t) - \phi_2(t)$ as a function of time for different values of the coupling constant c. For small coupling ($c = 0.1$) $\Delta\phi$ increases unbounded almost linearly in time, similar to the periodic case described by Adler's equation (19). If the coupling is increased above the critical value of $c \approx 0.18$ chaotic phase synchronisation occurs and $\Delta\phi$ undergoes a bounded chaotic oscillation.

This kind of phase synchronisation of chaotic oscillators [70] occurs also for large networks of coupled oscillators and may be viewed as a partial synchronisation (or coherence) because the amplitudes of the individual oscillators remain essentially uncorrelated. To synchronise their temporal evolution, too, stronger coupling is required and an (almost) perfect coincidence of all state variables of the coupled systems can, of course, be expected only if the systems are (almost) identical.

This kind of synchrony is called *identical synchronisation* and can be achieved by unidirectional coupling if some appropriate coupling scheme is used [66,67]. Furthermore, the driving chaotic system can by modulated by an external signal (a 'message') and synchronisation of the response system provides all information required to extract this signal from the (transmitted) coupling signal [67,71]. Whether (synchronising) chaotic systems are useful potential building blocks for secure communication systems is controversially discussed, because there are also powerful techniques from nonlinear time series analysis to attack such an encryption. If, for example, some part of the message input signal (plaintext) and the corresponding coupling signal (ciphertext) are known one may 'learn' the underlying relation induced by the (deterministic!) chaotic dynamics. An example for such a 'known plaintext attack' using cluster weighted modelling may be found in Ref. [54].

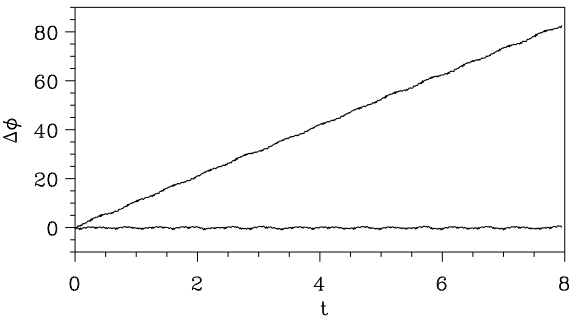


Figure 16. Phase difference $\Delta\phi = \phi_1 - \phi_2$ vs. time t for two representative cases: $c = 0.1$, $\Delta\phi$ grows linearly, no phase synchronisation; $c = 0.2$, $\Delta\phi$ is bounded, phase synchronisation [69].

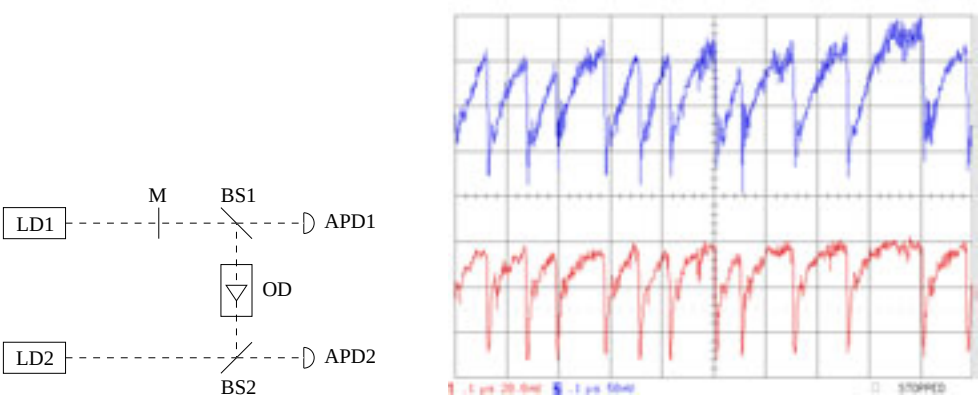


Figure 17. Experimental synchronisation of chaotic intensity fluctuations of two unidirectionally coupled semiconductor lasers. The lower trace (red) shows the irregular power drop outs of the drive laser and the upper trace (blue) the emitted light of the response laser [73].

5.3 Coupled semiconductor lasers with external cavities

As an example for (almost) identical synchronisation we show here two optically coupled semiconductor lasers. Each laser possesses an external resonator and the coupling is unidirectional due to an optical diode (Faraday isolator). Figure 17 show intensities of both lasers fluctuating in synchrony [19,72,73], a phenomenon we studied at the DPI with Volker Ahlers and Immo Wedekind.

5.4 Generalised synchronisation and parameter estimation

If the coupled systems are different from each other more sophisticated types of (generalised) synchronisation of chaotic dynamics may occur where asymptotically for $t \to \infty$ a function H exists that maps states of the driving system to those of the driven system [68,74–77].

An application of synchronisation of uni-directionally coupled systems, where generalised synchronisation plays an important role, is model validation and parameter estimation. Here, a measured time series drives a computer model and if the model is sufficiently accurate and all its parameters possess the right values one may achieve synchronisation between the computer model and the data. In this way, it is possible to recover those physical variables that have not been measured, as well as unknown parameters of the system. This is done by changing the parameters of the model until the (average) synchronisation error is minimized where generalised synchronisation is required to obtain a well defined (and smooth) error landscape. A practical example for this approach may be found in Ref. [78], where the parameters of a chaotic electronic circuit have been recovered, and improved methods are presented in Ref. [79].

6 Chaos Control

Since the end of the 1980 many chaos control methods have been suggested, studied, and applied [80–83] [8]. One of the most successful methods is time *Delayed Feedback Control* introduced by Pyragas [85] to stabilise unstable periodic orbits or fixed points embedded in a given chaotic attractor. For a general system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \tag{23}$$

with state vector $\mathbf{x}$ the control signal

$$\mathbf{u}(t) = k\left\{\mathbf{g}[\mathbf{x}(t-\tau)] - \mathbf{g}[\mathbf{x}(t)]\right\} \tag{24}$$

consists of the (amplified) difference between some observable $\mathbf{g}[\mathbf{x}(t)]$ and its time delayed value $\mathbf{g}[\mathbf{x}(t-\tau)]$. The parameter k is the gain of the feedback loop, $\mathbf{g}$ denotes a (suitable) measurement function and τ is the delay time, typically chosen to equal the period of the unstable periodic orbit (UPO) to be stabilised. In this case, the control signal vanishes on the periodic orbit, i. e. the UPO is not distorted by the control signal but only its stability properties are changed. Therefore, this kind of stabilisation is a *noninvasive* control method.

Delayed Feedback Control (DFC) is also called *Time Delay Autosynchronisation* (TDAS) and proved very useful for stabilising UPOs. However, it is less efficient for stabilising steady states (fixed points) because the control signal vanishes not only at the desired fixed point but for any τ-periodic solution. To impose a constraint that is fulfilled for constant solutions only, a second feedback term is necessary with a *different* delay time $\tilde{\tau}$ resulting in a control signal

$$\mathbf{u}(t) = k\left\{\mathbf{g}[\mathbf{x}(t-\tau)] - \mathbf{g}[\mathbf{x}(t)]\right\} + \tilde{k}\left\{\mathbf{g}[\mathbf{x}(t-\tilde{\tau})] - \mathbf{g}[\mathbf{x}(t)]\right\} . \tag{25}$$

If the ratio of delays $\tau/\tilde{\tau}$ is irrational, then there exists no periodic orbit on which the control signal vanishes. Only for fixed points $\mathbf{x}_0$ (with $\mathbf{g}[\mathbf{x}] = const.$) the differences in Eq. (25) vanish and the control signal equals zero, resulting in a noninvasive control. In general, more than two delay lines may be used and the gain factors of the delayed and the not delayed signals may be different. The control signal of such a *Multiple Delay Feedback Control* (MDFC) [86–89] is written as

$$\mathbf{u}(t) = k_0 + \sum_{m=1}^{M} k_{ma}\mathbf{g_m}[\mathbf{x}(t-\tau_m)] - k_{mb}\mathbf{g_m}[\mathbf{x}(t)] \tag{26}$$

with M different delay times τ_m and observables $\mathbf{g_m}$. For asymmetrical gains ($k_{ma} \neq k_{mb}$) this control signal provides in general an *invasive* control scheme but for steady state stabilisation the constant gain k_0 can be chosen in a way such that the control signal vanishes at the fixed point [89].

Multiple delay feedback control was introduced in collaboration with Alexander Ahlborn and turned out to be suffessful for controlling many dynamical systems [88]

[8] A general overview of control methods can be found in D. Guicking's article entitled 'Active control of sound and vibration' in this book [84]

with a clear tendency that control becomes the more effective the more different delay times are used. In the following two examples of successful application of MDFC are presented: stabilisation of a chaotic frequency-doubled Nd:YAG laser and manipulation of spatio-temporal dynamics of a complex Ginzburg-Landau equation.

6.1 Stabilisation of a frequency-doubled Nd:YAG laser

Multiple delay feedback control has been applied very successfully [87,90] to suppress chaotic intensity fluctuations of a compact frequency-doubled Nd:YAG laser which are notoriously difficult to avoid (green problem [91,92]). Figure 18(a) shows the experimental set-up where the laser's pump current provided by a current source is modulated by the feedback signal via a bias-T. The laser emits infrared radiation of wavelength 1064 nm as well as frequency-doubled green laser light with a wavelength of 532 nm. Both light emissions are separated from each other by a frequency selective beam splitter. Input signals of the control loop are either the two ac coupled orthogonally polarised infrared intensities $\tilde{I}_x$ and $\tilde{I}_y$ or the ac component of the green intensity $\tilde{G}$. Using the infrared signals the pump current modulation for MDFC may be written as

$$u(t) = a_x\tilde{I}_x(t-\tau_x) - b_x\tilde{I}_x(t) + a_y\tilde{I}_y(t-\tau_y) - b_y\tilde{I}_y(t) \tag{27}$$

and with the green laser light intensity

$$u(t) = a_x\tilde{G}(t-\tau_x) - b_x\tilde{G}(t) + a_y\tilde{G}(t-\tau_y) - b_y\tilde{G}(t)\,. \tag{28}$$

In both cases, the delay times τ_x and τ_y are typically in the range of $\tau_x \approx 0.6\,\mu$s and $\tau_y \approx 2.8\,\mu$s. All control parameters $a_x, b_x, a_y, b_y, \tau_x, \tau_y$ are chosen experimentally to achieve fixed point stabilisation. Fig. 18(b) shows a successful laser stabilisation using MDFC. Before the control signal is switched on at $t = 0$ s intensity fluctuations are visible which are then damped out by the feedback until the noise level is reached.

In this laser experiment three to four lasing modes were active. This case was also simulated [90] with an extended (multi mode) laser model describing an inhomogenous end-pumped YAG crystal.

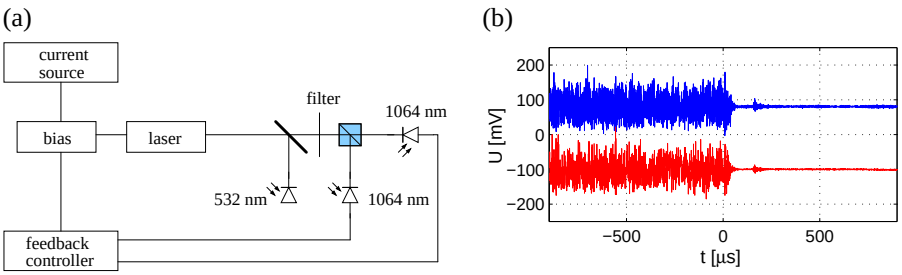


Figure 18. Suppression of chaotic intensity fluctuations of a frequency doubled Nd:YAG laser using MDFC as defined in Eq. (27). (a) Experimental setup. (b) Time series showing the orthogonally polarised ac coupled infrared signals $\tilde{I}_x$ (upper trace) and $\tilde{I}_y$ (lower trace). After activation of feedback control at $t = 0$ the chaotic fluctuations are suppressed.

Stabilisation of the Nd:YAG laser succeeded also with *Notch Filter Feedback* control [93] providing some easily implementable approximation of MDFC.

6.2 Controlling spatio-temporal chaos

Delayed feedback can also be used to locally stabilise and manipulate complex spatio-temporal dynamics [94]. To illustrate this approach we use the two-dimensional complex *Ginzburg-Landau equation* (GLE)

$$\partial_t f = (1+ia)\nabla^2 f + f - (1+ib)f|f|^2 + u \tag{29}$$

with an external control signal $u(\mathbf{x},t)$. ∂_t and ∇ denote the temporal and the spatial derivative, respectively. The GLE (29) is a prototypical equation for spatio-temporal systems close to a supercritical Hopf-bifurcation. It is solved here numerically for periodic boundary conditions with a spectral code based on a Runge-Kutta scheme of 4th order combined with a spectral method in space with a spatial grid of 90×90 elements ($\Delta x = \Delta y = 1$).

The control signal is applied at a few *control cells* $\mathcal{C}_i$, only, simulating experimental sensors and actuators. In general, the control signal u_i which is applied at cell $\mathcal{C}_i$

$$u_i(t) = \sum_{m=1}^{M} k_{i_{ma}} s_{i_j}(t - \tau_{i_m}) - k_{i_{mb}} s_{i_j}(t) \tag{30}$$

is given by delayed and non-delayed input signals s_{i_j} measured at other cells $\mathcal{C}_{i_j}$ where a measured signal

$$s_k(t) = \int_{\mathcal{C}_k} f(\mathbf{z},t)\,\mathrm{d}\mathbf{z} \tag{31}$$

is the averaged value of f at control cell $\mathcal{C}_k$.

Again, the performance of the control scheme depends crucially on the gains $k_{i_{ma}}, k_{i_{mb}}$ and delay times τ_{i_m} that may vary from cell to cell (as indicated by the multiple index).

Figure 19 shows two examples where different coupling schemes (Figs. 19(b) and 19(d)) are used to stabilise plane waves (Fig. 19(a)) and to trap a spiral wave (Fig. 19(c)).

7 Conclusion

This article is an attempt to give a tutorial overview of research in nonlinear dynamics at the DPI. Of course, it is incomplete but we hope it motivates the reader to learn more about this exciting interdisciplinary field that is heading now towards even more complex systems like large networks of coupled oscillators or swarms of interacting agents. So, stay tuned .... at DPI.

Acknowledgements. The author thanks Werner Lauterborn, Thomas Kurz, Robert Mettin and all other coworkers, students and staff at the DPI for excellent collaboration and support.

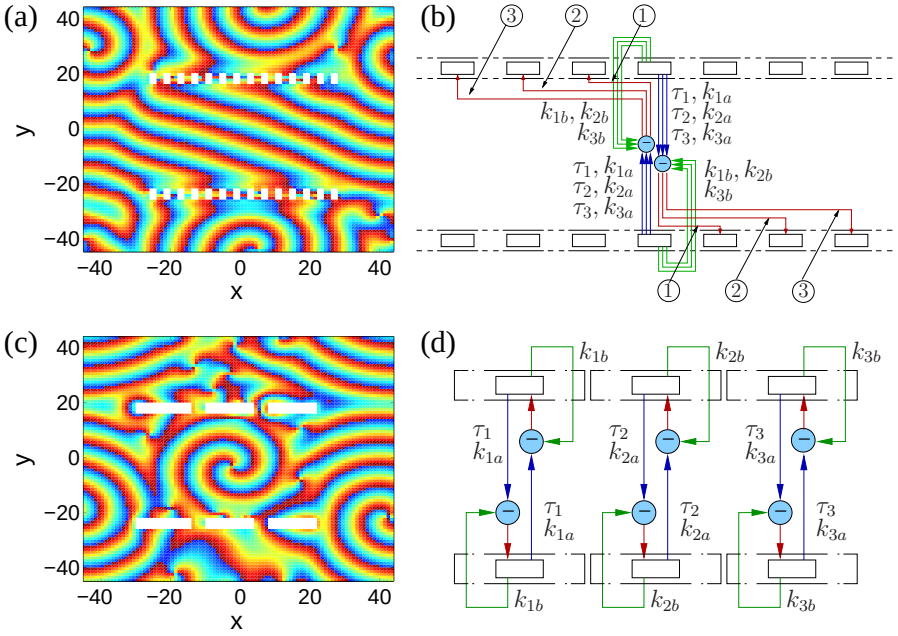


Figure 19. Phase (colour coded) of the complex solution f of the controlled GLE (29) with parameters $(a, b) = (-1.45, 0.34)$. White rectangles denote control cells where signals are measured and control is applied. In the region between the control cells chaotic spiral waves are turned into (a) slanted traveling waves if the control scheme (b) is applied with parameters $\tau_1 = 31$, $\tau_2 = 59$, $\tau_3 = 84$, $k_{1a} = 0.22$, $k_{1b} = 0.3$, $k_{2a} = 0.2$, $k_{2b} = 0.5$, $k_{3a} = 0.3$, and $k_{3b} = 0$. Using control scheme (d) with $k_{1a} = 0.22$, $k_{2a} = 0.1$, $k_{3a} = 0.35$, $k_{1b} = 0.3$, $k_{2b} = 0.5$, $k_{3b} = 0$, $\tau_1 = 41$, $\tau_2 = 27$, and $\tau_3 = 49$ individual spiral waves can be trapped (c). From Ref. [94].

References

[1] W. Lauterborn, 'Numerical Investigation of Nonlinear Oscillations of Gas Bubbles in Liquids', J. Acoust. Soc. Am. **59**, 283 (1976).

[2] W. Lauterborn and E. Cramer, 'Subharmonic Route to Chaos Observed in Acoustics', Phys. Rev. Lett. **47**, 1445 (1981).

[3] W. Lauterborn and U. Parlitz, 'Methods of chaos physics and their application to acoustics', J. Acoust. Soc. Am. **84**, 1975 (1988).

[4] G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*, (Verlag Friedr. Vieweg & Sohn, Braunschweig, 1918).

[5] U. Parlitz and W. Lauterborn, 'Superstructure in the bifurcation set of the Duffing equation $\ddot{x} + d\dot{x} + x + x^3 = f cos(wt)$', Phys. Lett. A **107**, 351 (1985).

[6] U. Parlitz and W. Lauterborn, 'Resonances and torsion numbers of driven dissipative nonlinear oscillators', Z. Naturforsch. **41a**, 605 (1986).

[7] C. Scheffczyk, U. Parlitz, T. Kurz, W. Knop and W. Lauterborn, 'Comparison of bifurcation structures of driven dissipative nonlinear oscillators', Phys. Rev. A **43**,

6495 (1991).
[8] R. Mettin, U. Parlitz, and W. Lauterborn, 'Bifurcation structure of the driven van der Pol oscillator', Int. J. Bifurcation Chaos **3**, 1529 (1993).
[9] U. Parlitz, 'Common dynamical features of periodically driven strictly dissipative oscillators', Int. J. Bifurcation Chaos **3**, 703 (1993).
[10] J. Ohtsubo, *Semiconductor Lasers - Stability, Instability and Chaos*, Springer Series in Optical Sciences (Springer Verlag, Berlin, 2006).
[11] C. Risch and C. Voumand, 'Self-pulsation in the output intensity and spectrum of GaAs-AlGaAs cw diode lasers coupled to a frequency-selective external optical cavity', J. App. Phys. **48**, 2083 (1977).
[12] R. Lang and K. Kobayashi, 'External optical feedback effects on semiconductor injection laser properties', IEEE J. Quantum Electron. **QE-16**, 347 (1980).
[13] M. Fujiwara, K. Kubota, R. Lang, 'Low-frequency intensity fluctuation in laser diodes with external optical feedback', Appl. Phys. Lett. **38**, 217 (1981).
[14] C. H. Henry and R. F. Kazarinov, 'Instability of semiconductor lasers due to optical feedback from distant reflectors', IEEE J. Quantum Electron. **22**, 294 (1986).
[15] A. Hohl, H. J. C. van der Linden, and R. Roy, 'Determinism and stochasticity of power-dropout events in semiconductor lasers with optical feedback', Opt. Lett. **20**, 2396 (1995).
[16] J. Mørk, B. Tromborg, and P. L. Christiansen, 'Bistability and low-frequency fluctuations in semiconductor lasers with optical feedback', IEEE J. Quantum Electron. **24**, 123–133 (1988).
[17] T. Sano, 'Antimode dynamics and chaotic itinerancy in the coherence collapse of semiconductor lasers with external feedback', Phys. Rev. A **50**, 2719 (1994).
[18] I. Fischer, G. H. M. van Tartwijk, A. M. Levine, W. Elsässer, E. Göbel, and D. Lenstra, 'Fast pulsing and chaotic itinerancy with a drift in the coherence collapse of semiconductor lasers', Phys. Rev. Lett. **76**, 220 (1996).
[19] V. Ahlers, U. Parlitz, and W. Lauterborn, 'Hyperchaotic dynamics and synchronization of external cavity semiconductor lasers', Phys. Rev. E **58**, 7208 (1998).
[20] M.-W. Pan, B.-P. Shi, and G. Gray, 'Semiconductor laser dynamics subject to strong optical feedback', Opt. Lett. **22**, 166 (1997).
[21] P. S. Spencer, and K. A. Shore, 'Multimode iterative analysis of the dynamic and noise properties of laser diodes subject to optical feedback', Quantum Semiclass. Opt. **9**, 819 (1997).
[22] I. Wedekind and U. Parlitz, 'Mode synchronization of external cavity semiconductor lasers', Int. J. of Bifurcation Chaos, in the press (2007).
[23] K. Peters, J. Worbs, U. Parlitz, and H.-P. Wiendahl, 'Manufacturing systems with restricted buffer size', in *Nonlinear Dynamics of Production Systems*, edited by G. Radons and R. Neugebauer (Wiley-VCH Verlag, Weinheim, 2004).
[24] K. Peters and U. Parlitz, 'Hybrid systems forming strange billiards', Int. J. Bifurcation Chaos **13**, 2575 (2003).
[25] A. Amann, K. Peters, U. Parlitz, A. Wacker, and E. Scholl, 'A hybrid model for chaotic front dynamics: From semiconductors to water tanks', Phys. Rev. Lett. **91**, 066601 (2003).
[26] P. Grassberger, and I. Procaccia, 'On the characterization of strange attractors', Phys. Rev. Lett. **50**, 346 (1983).
[27] R. Badii and A. Politi, 'Hausdorff dimension and uniformity factor of strange attractors', Phys. Rev. Lett **52**, 1661 (1984).
[28] R. Badii and A. Politi, 'Statistical description of chaotic attractors', J. Stat. Phys. **40**, 725 (1985).

[29] P. Grassberger, 'Generalizations of the Hausdorff dimension of fractal measures', Phys. Lett. A **107**, 101 (1985).
[30] C. Merkwirth, U. Parlitz, and W. Lauterborn, 'Fast Exact and Approximate Nearest Neighbor Searching for Nonlinear Signal Processing', Phys. Rev. E **62**, 2089 (2000).
[31] J. Holzfuss and G. Mayer-Kress, 'An approach to error-estimation in the application of dimension algorithms, in Ref. [32], pp. 114 (1986).
[32] G. Mayer-Kress (ed.), *Dimensions and Entropies in Chaotic Systems – Quantification of Complex Behavior*, (Springer, Berlin, 1986).
[33] J. Theiler, 'Estimating fractal dimension', J. Opt. Soc. Am. A **7**, 1055 (1990).
[34] G. Broggi, 'Evaluation of dimensions and entropies of chaotic systems', J. Opt. Soc. Am. B **5**, 1020 (1988).
[35] P. Grassberger, T. Schreiber, and C. Schaffrath, 'Nonlinear time sequence analysis', Int. J. Bifurcation Chaos **1**, 521 (1991).
[36] U. Parlitz, 'Nonlinear Time-Series Analysis', in *Nonlinear Modeling - Advanced Black-Box Techniques*, edited by J. A. K. Suykens and J. Vandewalle (Kluwer Academic Publishers, Boston, 1998), p. 209.
[37] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, (Cambridge University Press, Cambridge, 1997).
[38] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, (Springer, New York, 1996).
[39] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, 'The analysis of observed chaotic data in physical systems', Rev. Mod. Phys. **65**, 1331 (1993).
[40] H. D. I. Abarbanel and U. Parlitz, 'Nonlinear analysis of time series data', in *Handbook of Time Series Analysis*, edited by. B. Schelter, M. Winterhalder, and J. Timmer (WILEY-VCH Verlag, Weinheim, 2006), p. 5.
[41] K. Geist, U. Parlitz, and W. Lauterborn, 'Comparison of Different Methods for Computing Lyapunov Exponents', Prog. Theor. Phys. **83**, 875 (1990).
[42] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw,, 'Geometry from a time series', Phys. Rev. Lett. **45**, 712 (1980).
[43] F. Takens, 'Detecting strange attractors in turbulence', in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L. S. Young, (Springer, Berlin, Springer,1981), p. 366.
[44] T. Sauer, Y. Yorke, and M. Casdagli, 'Embedology', J. Stat. Phys. **65**, 579 (1991).
[45] T. Sauer and J.A. Yorke, 'How many delay coordinates do you need ?', Int. J. Bifurcation Chaos **3**, 737 (1993).
[46] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, 'State space reconstruction in the presence of noise', Physica D **51**, 52 (1991).
[47] J. F. Gibson, J. D. Farmer, M. Casdagli, and S. Eubank, 'An analytic approach to practical state space reconstruction', Physica D **57**, 1 (1992).
[48] D. S. Broomhead and G. P. King, 'Extracting qualitative dynamics from experimental data', Physica D **20**, 217 (1986).
[49] P. S. Landa and M. G. Rosenblum, 'Time series analysis for system identification and diagnostics', Physica D **48**, 232 (1991).
[50] M. Palus and I. Dvorak, 'Singular-value decomposition in attractor reconstruction: pitfalls and precautions', Physica D **55**, 221 (1992).
[51] T. Sauer, 'Reconstruction of dynamical systems from interspike intervals', Phys. Rev. Lett. **72**, 3811 (1994).
[52] R. Castro and T. Sauer, 'Correlation dimension of attractors through interspike intervals', Phys. Rev. E **55**, 287 (1997).
[53] D. M. Racicot and A. Longtin, 'Interspike interval attractors from chaotically driven neuron models', Physica D **104**, 184 (1997).

[54] D. Engster and U. Parlitz, 'Local and Cluster Weighted Modeling for Time Series Prediction', in *Handbook of Time Series Analysis*, edited by B. Schelter, M. Winterhalder, and J. Timmer (WILEY-VCH Verlag, Weinheim, 2006), p. 39.
[55] U. Parlitz and C. Merkwirth, 'Prediction of spatiotemporal time series based on reconstructed local states', Phys. Rev. Lett. **84**, 1890 (2000).
[56] G. Langer and U. Parlitz, 'Modelling parameter dependence from time series', Phys. Rev. E **70**, 056217 (2004).
[57] U. Parlitz, A. Hornstein, D. Engster, F. Al-Bender, V. Lampaert, T. Tjahjowidodo, S. D. Fassois, D. Rizos, C. X. Wong, K. Worden, and G. Manson, 'Identification of pre-sliding friction dynamics', Chaos **14**, 420 (2004).
[58] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization - A universal concept in nonlinear science* (Cambridge University Press, Cambridge UK, 2001).
[59] U. Parlitz, A. Pikovsky, M. Rosenblum, and J. Kurths, 'Schwingungen im Gleichtakt', Physik Journal **5**, Nr. 10, 33 (2006).
[60] M. Abel, S. Bergweiler, and R. Gerhard-Multhaupt, 'Synchronization of organ pipes: experimental observations and modeling', J. Acoust. Soc. Am. **119**, 2475 (2006).
[61] E. V. Appleton, 'The automatic synchronization of triode oscillator', Proc. Cambridge Phil. Soc. (Math. and Phys. Sci.) **21**, 231 (1922).
[62] B. van der Pol, 'Forced oscillations in a cicuit with non-linear resistance', Phil. Mag. **3**, 64 (1927).
[63] R. Adler, 'A study of locking phenomena inoscillators', Proc. IRE **34**, 351 (1946). Reprinted in: Proc. IEEE **61**, 1380 (1973).
[64] H. Fujisaka and T. Yamada, 'Stability theory of synchronized motion in coupled-oscillator systems', Prog. Theor. Phys. **69**, 32 (1983).
[65] A. S. Pikovsky, 'On the interaction of strange attractors', Z. Physik B **55**, 149 (1984).
[66] L. Pecora and T. Carroll, 'Synchronization in chaotic systems', Phys. Rev. Lett. **64**, 821 (1990).
[67] L. Kocarev and U. Parlitz, 'General approach for chaotic synchronization with applications to communication', Phys. Rev. Lett. **74**, 5028 (1995).
[68] U. Parlitz and L. Kocarev, 'Synchronization of Chaotic Systems', in *Handbook of Chaos Control*, edited by H. G. Schuster (Wiley-VCH Verlag, Weinheim, 1999), p. 271.
[69] U. Parlitz, L. Junge, W. Lauterborn, and L. Kocarev, 'Experimental observation of phase synchronization', Phys. Rev. E **54**, 2116 (1996).
[70] M. Rosenblum, A. Pikosvky, and J. Kurths, 'Phase synchronization of chaotic oscillators', Phys. Rev. Lett. **76**, 1804 (1996).
[71] U. Parlitz, L. Kocarev, T. Stojanovski, and H. Preckel, 'Encoding messages using chaotic synchronization', Phys. Rev. E **53**, 4351 (1996).
[72] I. Wedekind and U. Parlitz, 'Experimental observation of synchronization and anti-synchronization of chaotic low-frequency-fluctuations in external cavity semiconductor lasers', Int. J. Bifurcation Chaos **11**, 1141 (2001).
[73] I. Wedekind and U. Parlitz, 'Synchronization and antisynchronization of chaotic power drop-outs and jump-ups of coupled semiconductor lasers', Phys. Rev. E **66**, 026218 (2002).
[74] N. F. Rulkov,, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel, 'Generalized synchronization of chaos in directionally coupled chaotic systems', Phys. Rev. E **51**, 980 (1995).
[75] L. Kocarev and U. Parlitz, 'Generalized synchronization, predictability and equivalence of uni-directionally coupled dynamical systems', Phys. Rev. Lett. **76**, 1816 (1996).

[76] H. D. I. Abarbanel,, N. F. Rulkov, and M. M. Sushchik, 'Generalized synchronization of chaos: The auxiliary system approach', Phys. Rev. E **53**, 4528 (1996).
[77] U. Parlitz, L. Junge, and L. Kocarev, 'Subharmonic entrainment of unstable period orbits and generalized synchronization', Phys. Rev. Lett. **79**, 3158 (1997).
[78] U. Parlitz, L. Junge, and L. Kocarev, 'Synchronization based parameter estimation from time series', Phys. Rev. E **54**, 6253 (1996).
[79] W. Yu, G. Chen, J. Cao, J. Lü, and U. Parlitz, 'Parameter identification of dynamical systems from time series', Phys. Rev. E **75**, 067201 (2007).
[80] A. Hübler and E. Lüscher, 'Resonant stimulation and control of nonlinear oscillators', Naturwissenschaften **76**, 67 (1989).
[81] E. Ott, C. Grebogi, and J. Yorke, 'Controlling Chaos', Phys. Rev. Lett. **64**, 1196 (1991).
[82] *Handbook of Chaos Control*, edited by H. G. Schuster, (Wiley-VCH Verlag, Weinheim,1999).
[83] *Handbook on Chaos Control*, 2nd ed., edited by E. Schöll and H. G. Schuster (Wiley-VCH Verlag, Berlin, 2007).
[84] D. Guicking, 'Active Control of Sound and Vibration', in *Oscillations, Waves, and Interactions*, edited by T. Kurz, U. Parlitz, and U. Kaatze (Universitätsverlag Göttingen, Göttingen, 2007).
[85] K. Pyragas, 'Continuous control of chaos by self-controlling feedback', Phys. Lett. A **170**, 421 (1992).
[86] A. Ahlborn and U. Parlitz, 'Chaos control of an intracavity frequency-doubled Nd:YAG laser', Proceedings of the Experimental Chaos Conference 8 (ECC8), Florence 14.6.–17.6.2004, AIP Conf. Proc. **742**, 241 (2004).
[87] A. Ahlborn and U. Parlitz, 'Stabilizing Unstable Steady States Using Multiple Delay Feedback Control' Phys. Rev. Lett. **93**, 264101 (2004).
[88] A. Ahlborn and U. Parlitz, 'Controlling dynamical systems using multiple delay feedback control', Phys. Rev. E **72**, 016206 (2005).
[89] A. Ahlborn and U. Parlitz, 'Multiple Delay Feedback Control', in *Handbook on Chaos Control*, 2nd ed., edited by E. Schöll and H. G. Schuster (Wiley-VCH Verlag Berlin, 2007).
[90] A. Ahlborn and U. Parlitz, 'Laser stabilisation with multiple-delay feedback control', Opt. Lett. **31**, 465 (2006).
[91] T. Baer, 'Large amplitude fluctuations due to longitudinal mode-coupling in diode-pumped intra- cavity-doubled Nd:YAG lasers', J. Opt. Soc. Am. B **3**, 1175 (1986).
[92] A. Schenck zu Schweinsberg and U. Dressler, 'Characterization and stabilization of the unstable fixed points of a frequency doubled Nd:YAG laser', Phys. Rev. E **63**, 056210 (2001).
[93] A. Ahlborn and U. Parlitz, 'Chaos Control using Notch Filter Feedback', Phys. Rev. Lett. **96**, 034102 (2006).
[94] A. Ahlborn and U. Parlitz, 'Controlling spatiotemporal chaos using multiple delays', Phys. Rev. E **75**, 065202(R) (2007).

Oscillations, Waves and Interactions, pp. 435–460
edited by T. Kurz, U. Parlitz, and U. Kaatze
Universitätsverlag Göttingen (2007) ISBN 978–3–938616–96–3
urn:nbn:de:gbv:7-verlag-1-16-3

DPI60plus – a future with biophysics

S. Lakämper and C. F. Schmidt

Drittes Physikalisches Institut, Georg-August-Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

Abstract. In this review we first give a short introduction into the techniques currently in use and development to establish biophysics as a field of research at the DPI. On this basis, we then continue to sketch recent research highlights, covering the growing group's entire scientific range. Examples are presented to illustrate the intriguing physical complexity of biological matter and the wealth of physical approaches to study it.

The research focus of the Drittes Physikalisches Institut is changing with a change of guards in 2006. The central activities will be in the area of biophysics and physics of complex systems. Biophysics is an interdisciplinary and rather broad field of research, with strong ties to condensed matter physics, statistical physics and various kinds of technical branches of physics. Here we want to highlight recent advances in a variety of projects in the biophysics group. We want to show how approaches and techniques from physics can help to understand very diverse systems from single molecules to complex polymer-networks in soft condensed matter and artificial cell-systems, as well as real cells and tissues. This overview is not intended to present a complete review of the field, but rather to provide a snapshot of current activities.

Experimental research hinges on technologies, and to be on the cutting edge often requires the development of new approaches that can open new fields of inquiry. We use and further develop a variety of approaches, grouped around so called single molecule techniques such as Atomic Force Microscopy (AFM), single-molecule fluorescence microscopy, optical trapping techniques and combinations thereof. The following sections give an introduction to these methods before we touch on current research projects.

1 Introduction to technologies

1.1 Atomic force microscopy

AFM – developed in the 1980s by Binnig and Rohrer as well as Quate and Hansma – initially as an expansion of Scanning Tunneling Microscopy (STM) – has evolved into an important research tool, particularly in biophysics. AFM probes surfaces by mechanical scanning with a nanometre-sized sharp tip mounted to a pliant cantilever. The deflection of a laser beam reports the force exerted on the tip, which is converted to a topographic image of the surface after 2D-scanning the object of

interest. In contrast to conventional microscopy, the AFM reports the response of the imaged objects to the force exerted by the tip which can give more than just structural information about the sample. While generally rather slow compared to, e. g., video microscopy, AFM provides nanometre or better resolution and therefore can resolve details of bio-macromolecules that are otherwise only accessible by electron microscopy or X-ray crystallography.

The AFM furthermore allows one to measure forces with piconewton resolution. This capability can be used to mechanically probe single molecules, but also biopolymers, such as DNA or protein chains, and determine rigidity, rupture forces or unfolding forces. AFM-imaging of biomolecules is generally still a slow technique, requiring 10 s or more of seconds per frame. We are especially interested in new developments aiming for imaging at video rates in order to capitalize on the capability of AFM to monitor structure, mechanics and dynamics at the same time in physiological conditions. AFM can also be combined with fluorescence microscopy which adds specific recognition.

1.2 Fluorescence microscopy

Light microscopy and especially fluorescence microscopy has experienced a renaissance with the advent of laser illumination and highly photo-stable chemical and biological fluorophores. We are especially focussing on the imaging of single molecules. The development of specific labelling strategies and of highly sensitive detection methods and cameras have made the real-time observation of single molecules – *in vitro* – or even in living organisms – *in vivo* – possible.

To be able to detect single molecules, the background fluorescence needs to be sufficiently low. Two approaches are used in the lab: using total internal reflection of a laser-beam on the glass–water interface of the sample chamber, the (evanescent wave) illumination-depth within the sample is reduced to 100–200 nm. This drastically reduces the background fluorescence, as the typically tens of μm-thick samples are not completely illuminated. The other approach is to use wide-field illumination and strongly reduce the concentration of active fluorescent proteins. The latter approach allows better control over illumination intensities and the bleaching processes. Furthermore, we are currently developing multi-colour single molecule fluorescence setups to use Förster-Resonance Energy Transfer (FRET) between pairs of fluorophores. Since the energy transfer is strongly dependent on distance, it can be used to monitor domain and/or sub-unit interactions of proteins (molecular motors/chaperonins) on the nanometre scale.

Diffraction-limited imaging of single fluorophores results in diffraction patterns of size $\sim \lambda/2$ which limits the spatial resolution in densely labeled samples. An individual fluorophore can, however, be localized by fitting the diffraction pattern of a point source with accuracies better than 2 nm, which provides valuable information about the dynamics of molecular machines. The accuracy of position detection and relative shifts can be combined with or complemented by the sub-nm resolution of optical trapping techniques.

1.3 Optical trapping

Optical trapping exploits the transfer of momentum due to scattering or refraction of photons by refractile objects. The forces on a small particle of higher index of refraction than its surroundings (for example a latex or glass bead in water) can be made to trap the particle near the focus of the laser beam. Several aspects make such an optical trap (or "tweezers") particularly interesting for the study of single biomolecules: 1.) the force-range accessible with optical traps is – dependent on laser power – about 0–250 pN which well matches the forces generated by individual motor proteins and thus fills the gap between load-free conditions in fluorescence experiments and the minimal forces resolvable by AFM (> 50 pN); 2.) the Brownian motion of the bead in the trapping potential is well measurable and can thus be used to report binding and unbinding of individual molecules to their substrate. Binding means additional spatial confinement or an increase in total system stiffness which results in a decrease of displacement variance. Such measurements can be performed with a time resolution of 1 ms which is sufficiently high for the study of many conformational changes in proteins, for example motor proteins. 3.) The spatial resolution in optical trapping set-ups using interferometric detection is equally well suited for conformational changes of many biomolecules, namely in the nanometre range. Acusto-optical deflectors make it possible to rapidly steer the trap, either to create a time-dependent force on molecules or to switch between multiple quasi-simultaneous trap positions.

1.4 Microrheology

Currently, optical traps are, on the one hand, used in the lab to measure the forces and the steps molecular motors produce when they move along cytoskeletal filaments. Optical trapping and fast and accurate position detection are, on the other hand, also used for "microrheology", i. e. to probe the dynamic viscoelastic properties of soft systems such as colloidal suspensions or polymer networks on mesoscopic scales. Soft materials are important in technology. Examples are plastics, synthetic polymers, polymer solutions, colloids and gels. Most biomaterials also classify as soft materials, such as cytoskeletal protein polymers, polysaccharides, lipid membranes or DNA solutions. Many of the varied and intriguing properties of soft materials stem from their complex structures and dynamics with multiple characteristic length and time scales. One of the characteristic and frequently studied material properties of such systems is their shear modulus. In contrast to ordinary solids, the shear modulus of polymeric materials can exhibit significant time or frequency dependence in the range of milliseconds to seconds or even minutes. In fact, such materials are typically *viscoelastic*, exhibiting both a viscous and an elastic response.

Rheology, which is the experimental and theoretical study of viscoelasticity in such systems, is of both fundamental and immense practical significance. Bulk viscoelasticity is usually measured with mechanical rheometers that probe macroscopic milliliter samples at frequencies up to tens of Hertz. Recently, a number of techniques have been developed to probe the material properties of systems ranging from polymer solutions to the interior of living cells on microscopic scales. These techniques have come to be called *microrheology*, as they can be used to locally measure viscoelastic

parameters. There have been several motivations for such developments. In many cases, and especially in biological systems, samples only come in small sizes. Another strong motivation for biological applications has been the prospect of being able to study inhomogeneities, for instance inside of cells. Furthermore, such techniques have provided the possibility to study viscoelasticity at frequencies far above 1 kHz. Finally, the ability to study materials such as polymer solutions with probes spanning some of the characteristic microscopic length scales, e. g. approaching the inter-chain separation or mesh size of gels, has led to new insights into the microscopic basis of viscoelasticity in these systems.

We use these techniques to measure the frequency dependence of the shear elastic modulus of both technical polymers and colloids, biological filamentous networks and even whole cells. We use several different experimental approaches: in passive microrheology we merely monitor either the fluctuations of individual probe particles (one-particle passive microrheology) or the correlated fluctuations of pairs of particles (two-particle passive microrheology). In active microrheology we exert oscillating forces on one bead with the help of the trap and AODs and monitor the response of a second particle. Cytoskeletal networks, for example entangled or cross-linked actin networks, have been a focus of interest. *In vitro* reconstituted networks are a step towards the highly complex and multi-component cytoskeleton of cells. An important step in the direction of the real systems is the addition of molecular motors to such model networks. Myosin motors can interact cyclically with the actin filaments under ATP consumption and create tension in the network. In this situation the system is out of equilibrium. The understanding of such non-equilibrium systems is of value to the understanding of cellular systems which are almost by definition out of equilibrium. A next step in complexity is to couple such non-equilibrium networks to uni-lamellar lipid vesicles. Such systems are also a step on the way to an artificial cell. In complementary approaches we also optically manipulate particles attached to or introduced into living cells.

2 Biomolecular shell mechanics probed with atomic force microscopy

2.1 Microtubules

Microtubules, one of the three major types of cytoskeletal protein-filaments are polarized polymers of tubulin. The 25 nm-diameter hollow tubules not only provide a mechanical scaffolding for eukaryotic cells, but also form tracks for motor proteins (kinesins and dyneins) which move various cargoes in a preferential direction along the microtubules. One of our recent studies aimed at high-resolution imaging of the nm-spacing of the tubulin subunits in the microtubule lattice. As can be seen in Fig. 1 – imaging resolves the building blocks of the microtubules and shows a distinct difference in the topography of the interior and exterior surfaces of microtubules: the exterior shows a clear radial periodicity of about 5 nm, corresponding to the spacing of the protofilaments, while the interior surface reveals also the axial spacing of tubulin subunits, reflected in a distinct 4 nm repeat [1].

The AFM tip can readily image the subunit structure when the forces used for imaging are well controlled and low enough (~ 100 pN), given the limited stability

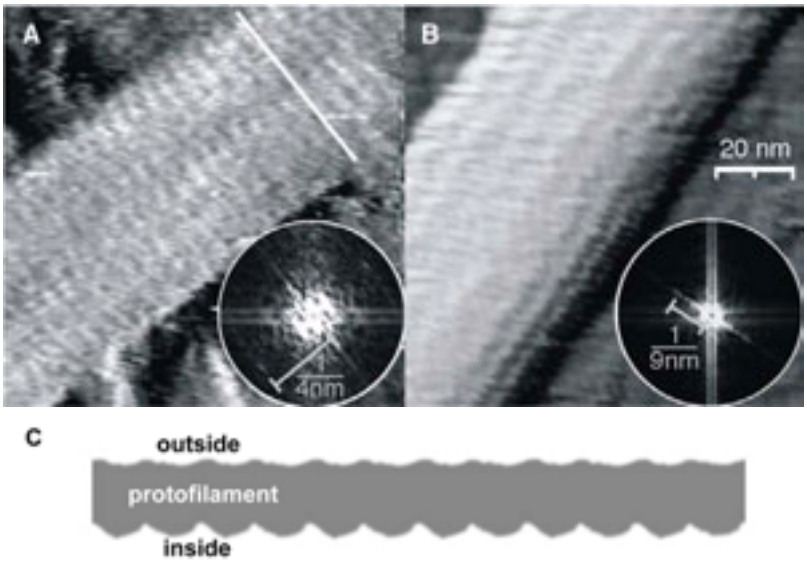


Figure 1. Both scans show a 100 nm × 100 nm region scanned with 128 × 128 points with a maximum tip force of 100 pN. A derivative filter was applied. (A) Opened microtubule (MT) on a DETA surface showing the inner surface of the wall; the protofilaments are hardly visible, but a striated pattern is visible oriented roughly at a right angle to the MT axis. The inserted line is exactly perpendicular to the MT axis, showing the angle of the stripes. A fast-Fourier-transformed image (FFT) (inset) shows weak peaks corresponding to a periodicity of 4 nm, the size of a tubulin monomer. (B) Intact MT on an APTS surface showing the outer surface imaged under similar conditions. The protofilaments are visible. Both in the topography as well as in the FFT there is no indication of the axial monomer periodicity. The protofilaments give a visible, but not very clear, signature in the FFT, because only five are visible and their apparent spacing is not constant. (C) Sketch of the axial cross-section of a protofilament based on cryo-EM results by Nogales et al. (1999). The periodicity of the monomers is much more pronounced on the inside. This is consistent with the finding that only an opened MT shows monomer periodicity in the axial direction (see (A)) (from Ref. [1]).

of protein–protein interactions. A fourfold increase in force (to 400 pN) results in microtubule destruction [1]. Forces at the limit of destruction occasionally do not result in complete destruction of the microtubule, but cause local defects as shown in Fig. 2. Such defects can span several or only 1–2 tubulin subunits. Repeated imaging of the defective area at low force revealed unequivocally that such defects can anneal, i. e. the tubulin subunits are able to rearrange such that the gap is closed [1,2].

2.2 Microtubule associated proteins and their influence on microtubule stability

The success of imaging intact microtubules in physiological solutions sparked the idea to image not only the microtubule itself but also microtubule associated proteins (MAPs; e. g. tau and double cortin) and molecular motors (kinesin). The MAP

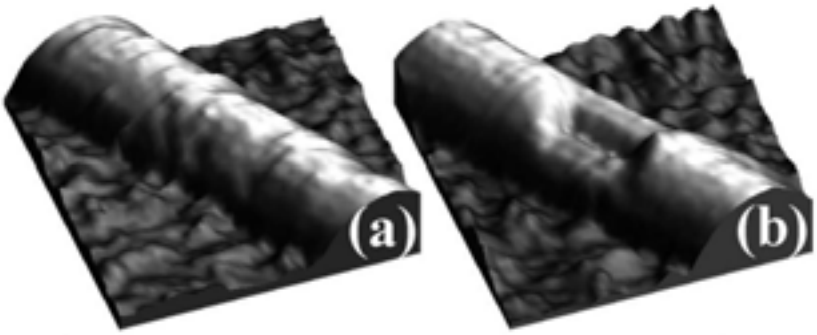


Figure 2. (a) shows a typical SFM image of an MT before performing a set of force-distance measurements (FZ). In (b) a hole can be observed at the spot of the microtubule where the FZs where performed after the detection of catastrophic breakages in the force distance curves (from Ref. [2]).

tau is one of the most abundant microtubule associated proteins and is involved in the stabilization and bundling of axonal microtubules in neurons. Tau is also infamous as a major component of the fibrillar structures correlated with human neurodegenerative diseases such as Alzheimer's. Although intense research has revealed much about tau function and its involvement in Alzheimer's disease, it has remained unclear how exactly tau binds to microtubules [3].

In a recent study we used AFM to image microtubules at saturating tau concentrations and found an increase in diameter of tau-decorated microtubules of 2 nm. While tau slightly increased the damage threshold of microtubules, measuring the radial stiffness of decorated microtubules revealed no difference to undecorated microtubules. Together with the finding that tau binding leaves the proto-filament structure well visible, this finding is consistent with the model that tau binds along the ridge of a proto-filament. Finite-element modelling confirmed that the radial elasticity should be unaffected by tau decoration in that way [3,4].

In contrast to tau, the MAP doublecortin (DCX) has been reported to bind on the outside of microtubules between the protofilaments. Finite-element modelling of that geometry predicts an increased radial stiffness of decorated microtubules. DCX has been found to be of importance for neuronal development. DCX mutations lead to mislocalization of nuclei in developing neurons and DCX dysfunction in humans leads to the disorder lisenzephaly. Ongoing AFM experiments with DCX-decorated microtubules have not yet shown a substantially increased radial stiffness.

2.3 Imaging motor proteins using AFM

We are also investigating the movement of motor proteins on microtubules by AFM. While dynein is a rather large roughly globular protein and should therefore be well suited for visualization by AFM, it is very difficult to prepare in pure and active form. Furthermore the flexibility of the dynein stalk might make imaging of dynein challenging. Kinesin motors, on the other hand, bind tightly and are relatively easily

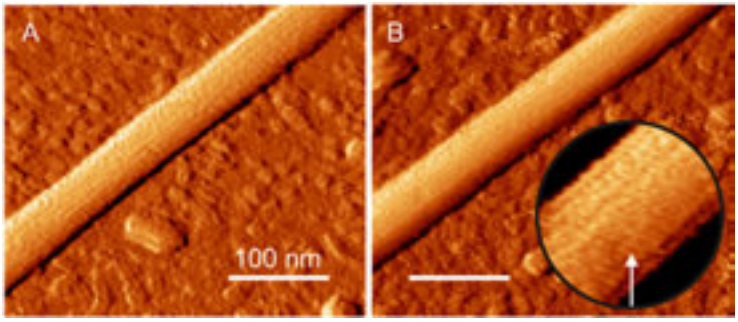


Figure 3. AFM scans of MTs. Scale bars represent 100 nm. Because of tip-sample dilation the MT width appears exaggerated (Schaap et al., 2004, Ref. [1]). (A) MT without tau, showing clearly the protofilaments. (B) For MTs with tau (ratio of 1:1 of tau:tubulin monomers), the protofilaments are still visible. The height increased by 2 nm (see also Fig. 4). Inset: this zoom shows a loose fibre with a height of ~ 0.5 nm that could occasionally be seen. (From Schaap et al., J. Struct. Biol., 2007, Ref. [3]).

prepared. They are similar in size to tubulin subunits and move with velocities of up to 1 μm/s along the microtubule at saturating ATP conentrations. We succeeded in imaging microtubules fully decorated with kinesin and measured a significant increase in diameter (see Fig. 4). We also observed clusters and single kinesin motors on microtubules. Repeated scanning indicates that we are able to follow individual kinesin motors moving along the microtubule. The technical challenge is to increase the rate of AFM-imaging to video-rate. The biophysics group will focus on the development of fast AFMs for this and other applications.

2.4 Viral capsids

We have further applied AFM to study the structure and mechanical properties of viral capsids. The particular viruses we have studied are the bacteriophage Phi29 and the plant virus cowpea chlorotic mottle virus (CCMV). Much like microtubules, viral capsids are self-assembling structures with typical sizes of tens of nanometres. Most viral capsids have highly regular and symmetric structures of more or less icosahedral symmetry. The shells are assembled from a well defined number of copies of mostly just one structural protein. Packaging of the DNA into the capsid is in the case of bacteriophages driven by motor proteins to such packing density that the capsid has to resist considerable outward directed forces, translated to a pressure about 60 atm [5].

Phi29 capsids deformed elastically under the AFM tip up to a force of about 1 nN and we could model the initial linear response of the shells by a simple homogeneous shell model. Under higher forces the shells fractured and collapsed. CCMV virus shells have the particular property that they expand under a change of pH. We observed that this expansion which goes along with an effective thinning of the shell wall caused a transition between two very different elastic responses. At low pH, in the condensed state, shells deform linearly and then buckle and fracture, whereas in

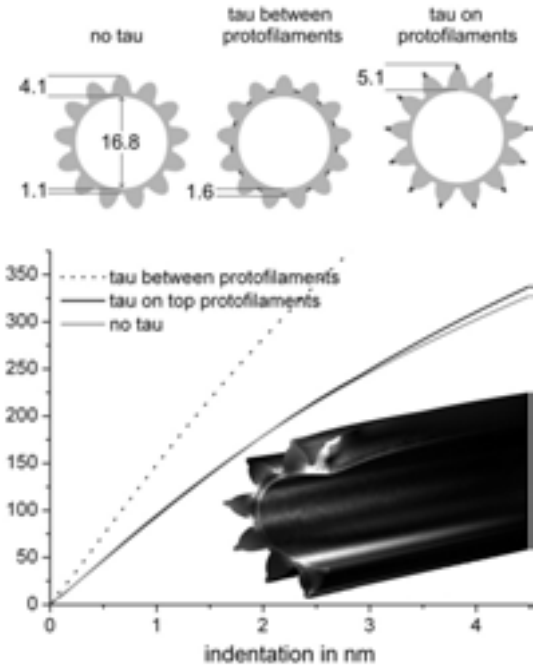


Figure 4. Finite-element simulations of MTs with bound tau. On top, the cross-sections of the models with their dimensions in nm are given. Left, MT in the absence of tau. Center, tau is added as a 0.5 nm thick layer between the protofilaments. Right, tau is added as 1 nm thick filaments on the ridges of the protofilaments. For all models the elastic modulus of the added material was set to 0.6 GPa, which is equal to that of the MT. The graph shows the computed deformation of the tube when indented with a parabolic tip with a 20 nm radius. The addition of tau as 1 nm filaments on top of the proto-filaments adds very little to the probed stiffness, but when tau is added as a 0.5 nm layer between the protofilaments, the stiffness increases by more than 60 %. The inset shows the MT with tau on top of the proto-filaments. The strain (indicated by brighter colors) is concentrated at the loading point and between the proto-filaments. (From Ref. [3]).

the expanded state they become super-elastic and can be reversibly compressed until the opposing walls touch [5–7].

2.5 Self-assembled DNA-tetrahedra

The smallest structures we have studied are nanometre-sized cages of DNA the mechanics of which are similar to those of the shells described above. DNA is an interesting material for the construction of nanomaterials because its self-assembly can be pre-programmed by the sequence of bases. With the tools of current molecular biology DNA oligomers can be generated with any desired sequence. Together with a group in Oxford, we have studied a family of DNA nanostructures that were designed to self-assemble to tetrahedra with double-stranded edges in a single step in only a

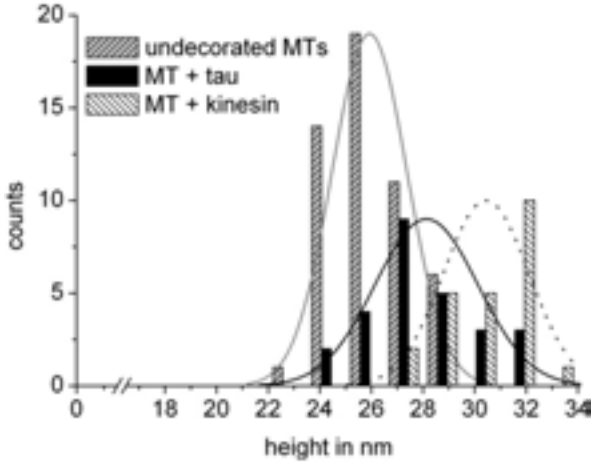


Figure 5. Height distribution of MTs. Undecorated MTs show an average diameter of 25.9 nm at a loading force of < 0.1 nN. When MTs are preincubated with tau the height increased by 2.2 nm. Kinesin added 4.5 nm to the MT height. Note that the 2.2 nm are caused by tau binding all around MTs while kinesin is bound to MTs just on one side because in this case MTs were attached to the surface before addition of kinesin. The differences between the three distributions were tested for statistical significance by a Student's t-test. The probability that any pair of the measured distributions belong to the same parent distribution was estimated to be smaller than 0.1 % by applying the Student's t-test for samples with an unequal variance. (From Ref. [1]).

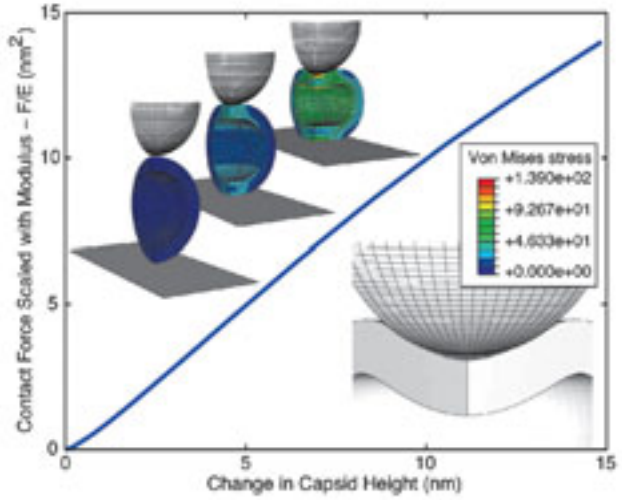


Figure 6. Calculated dependence of F/E on CCMV-capsid deformation for $R = 14.3$ nm, $h = 3.8$ nm, and $s = 0.4$. The images above the curve show a one-quarter segment of the capsid at indentations $d = 0$, 5.6, and 14 nm, and the von Mises stress is indicated by the colour. The image below the curve shows the buckling of the capsid away from the tip that is mirrored by a decrease in the slope. (From Ref. [5]).

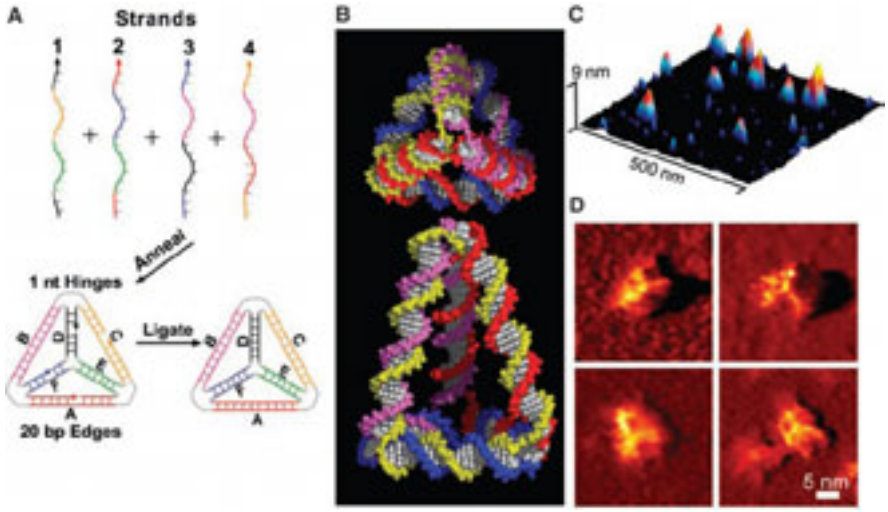


Figure 7. DNA tetrahedra. (A) Design of a DNA tetrahedron formed by annealing four oligonucleotides. Complementary subsequences that hybridize to form each edge are identified by colour. (B) Two views of a spacefilling representation of a $3 \times 20/3 \times 30$-bp tetrahedron. The backbone of each oligonucleotide is indicated by a single colour. (C) AFM image showing several tetrahedra on a mica surface. (D) AFM images, recorded with ultrasharp tips, of four tetrahedra; the three upper edges are resolved. (From Ref. [8]).

few seconds [8]. The desired structures can be generated with yields as high as 95%. We have demonstrated the versatility of this recipe to generate building blocks for 3D nanofabrication by assembling one regular and nine different irregular tetrahedra and by connecting them with programmable DNA linkers. The DNA tetrahedra are designed to be mechanically robust; they consist of rigid triangles of DNA helices covalently joined at the vertices (Fig. 7(A)). The four component oligonucleotides each run around one face and hybridize to form the doublehelical edges. We have used AFM to image the tertiary structure of individual tetrahedra and to demonstrate their rigidity, which we have then exploited to measure the response of DNA to axial compression. The triangulated stable construction of the tetrahedra is the only geometry in which compressional deformation of DNA has ever been achieved in a controlled way.

The tetrahedra imaged by AFM in Fig. 7(C) and (D), were designed to have three 30-base pair (bp) edges meeting at one vertex and three 20-bp edges bounding the opposite face (a molecular model is shown in Fig. 7(B)). They are expected to bind to a surface in one of two orientations, with heights of 10.5 nm if resting on the small face and 7.5 nm if resting on any of the other three faces. Figure 7(C), recorded with a tip 20 nm in radius, shows several objects with heights consistent with the two orientations. Figure 7(D) shows high-resolution images, obtained using ultra-sharp tips with radii of only 2 to 3 nm, that resolve the three upper edges of individual tetrahedra [8].

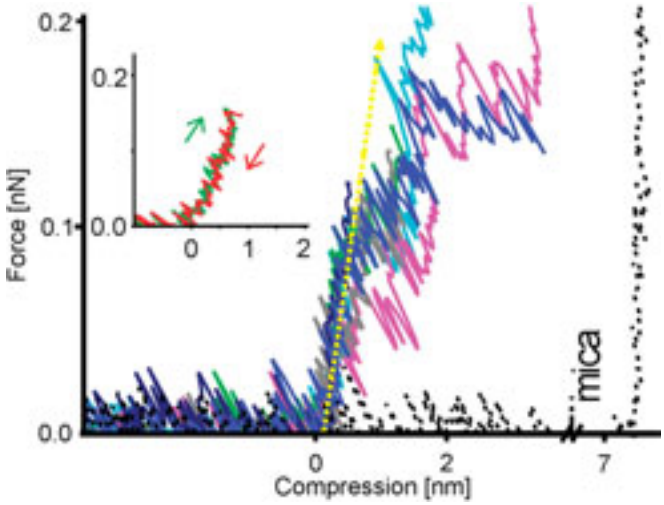


Figure 8. Compression of single DNA tetrahedra. Compression curves show linear elastic response up to a load of 0.1 nN. At higher forces, most tetrahedra deform irreversibly. Offsets were adjusted to overlap the linear parts of the seven curves. Inset: Reversibility of the elastic response of a typical tetrahedron. (Figure from Goodman et al., Science 2005, Ref. [8]).

We used these structurally braced tetrahedra to investigate the behaviour of DNA under compression. Although DNA under tension has been widely studied [9–12], DNA strands of micrometer length buckle at extremely low forces. To measure the mechanical response of a single tetrahedron directly, the AFM tip was centered over a tetrahedron first located in imaging mode and was then moved toward the surface while recording force. Compression curves for seven distinct $3 \times 20/3 \times 30$-bp tetrahedra are shown in Fig. 8. For forces up to 100 pN, the response was approximately linear and reversible (Fig. 8, inset) with an average force constant of $0.18(\pm 0.07)\,\mathrm{Nm}^{-1}$. At higher forces, the response was nonlinear and varied from tetrahedron to tetrahedron. From the gradient of the linear part of the measured F-d curve, we infer an elastic modulus of $K_c = 0.7(\pm 0.3)$ nN for one DNA double helix in compression [8].

3 Motor proteins studied by single-molecule fluorescence and optical trapping

3.1 Introduction: Kinesin function and structure

Molecular proteins are enzymes which use ATP-hydrolysis to move cargoes along cytoskeletal filaments (microtubules or actin-filaments). We study how members of the kinesin family of motor proteins function in microscopic detail. The kinesin family consists of several classes of motors which are structurally and kinetically diverse, but share high similarity in the motor domain (or head), which is responsible for ATP- and microtubule binding. In the case of conventional kinesin (Kinesin-1) two identical kinesin heavy-chains (KHC) – bearing the head-domain on the N-terminal

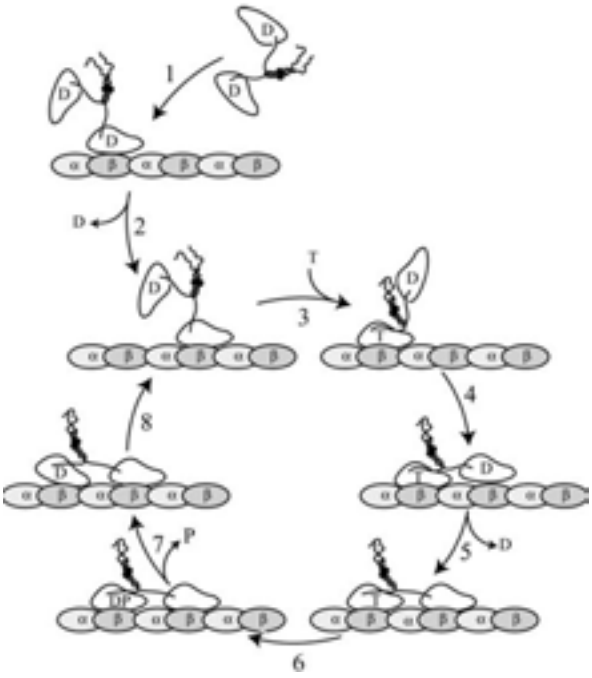


Figure 9. Schematic working hypothesis for the processive movement of Kinesin-1 Motors. For details see text. (from Ref. [14]).

end – dimerize via an extended alpha-helical "coiled-coil" of about 60 nm length. This rather rigid structure is interrupted by short stretches of disordered domains which allow the dimer to fold internally, such that the most C-terminal domain, the tail, and its bound kinesin light chains (KLC) can interact with the head and initial stalk (also termed neck). This interaction serves as an internal energy saving mechanism, as cargo also attaches through adapter molecules to the KLC and tail domains: when there is no cargo bound the motor self-inhibits by back-folding to the head and thus prevents futile ATP-consumption in the cell [13].

Kinesin-1 binds and transports cargoes such as vesicles over long distances, for example through axons of nerve cells which can, in extreme cases, be 1 m in length. Kinesin-1 dimers have structurally evolved to be able to bind to microtubules in a cyclical, nucleotide-dependent manner such that one head remains bound to the microtubule at any given time. This "processivity" is terminated in a statistical manner (Poisson process), on average after about 100–150 cycles. With a spacing of 8 nm between kinesin binding sites on the microtubule lattice, a single motor dimer can generate up to 7 pN force. Processivity ensures that cargo can be transported by few motor molecules. Through single-molecule and biochemical assays a basic model for Kinesin-1 stepping has emerged.

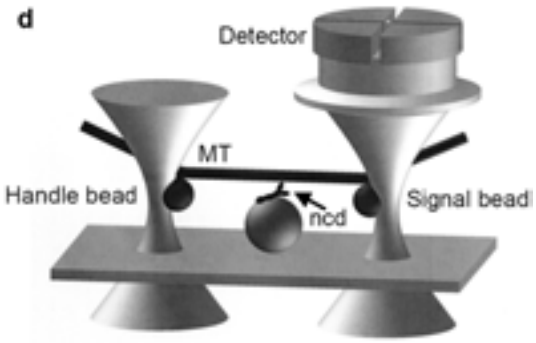


Figure 10. Schematic representation of the three bead-assay used for the determination of non-processive motor interactions (from Ref. [16]).

3.2 The chemo-mechanical cycle of Kinesin-1

According to this model, Kinesin-1 dimers in solution have ADP bound and can only bind weakly to microtubules. When a motor interacts with the microtubule, one of the two heads rapidly looses the ADP, forming a tightly bound, nucleotide-free state while the second head is prevented from binding to the microtubule until the nucleotide free, bound head binds ATP from solution. ATP-binding loosens the neck-linker such that the second head can bind to the microtubule and release the bound ADP [13–15]. When the hydrolysis product phosphate is released, the rear head dissociates and the cycle starts again. Fluorescence microscopy allows us to visualize the movement of single Kinesin-1 dimers under load-free conditions, and optical trapping experiments make it possible to exert and measure forces.

3.3 Kinesins – structural adaptations for specific cellular tasks

Kinesins come in a variety of forms and functions, and not all are processive. We also study non-processive kinesins which employ a myosin-like conformational change of a lever-arm. One example is the dimeric Kinesin-14 ncd. This kinesin plays a role in the formation of the meiotic and mitotic spindle. It is a minus-end-directed motor, and kinetic and structural studies indicate that ncd dimers do not move processively along a microtubule. We have detected the power stroke of ncd using a double-laser trap and a so-called three bead assay. In the three bead assay, ncd was adsorbed onto a large ($5\,\mu$m) bead fixed to a coverslip. A microtubule was then suspended and manipulated above the stationary bead with a dual-beam laser trap (see Fig. 10, Ref. [17]). Figure 11 shows a typical time trace of the bead positions, and the corresponding standard deviation from which binding events can be detected. Because the data were noisy, the power stroke of about 9 nm could only be extracted from ensemble-averaged events [16].

Kinesins differ not only in their processivity characteristics, but also in structure. Some kinesins are monomeric and some are trimeric or tetrameric, adapted for the specific roles in the cell. We currently investigate the motile behaviour of a

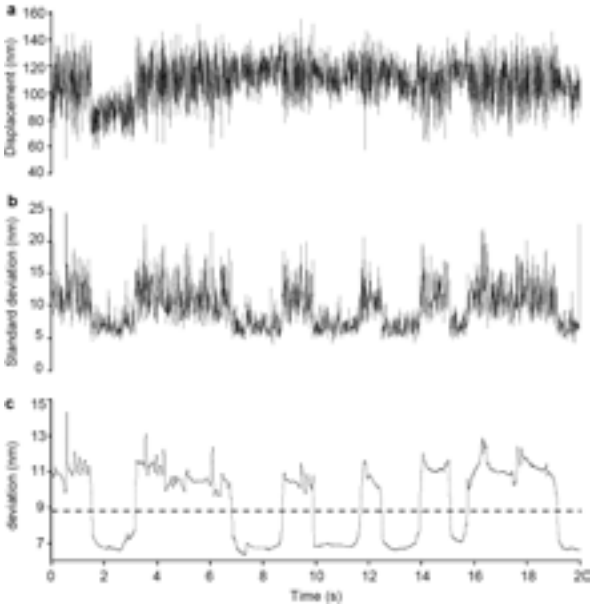


Figure 11. Single-molecule binding events generated by full-length ncd in a three-bead assay. (a) Raw data, 2-kHz sample rate, immobilized ncd interacting with a suspended microtubule in the presence of 2 μM ATP. Displacements are measured parallel to the long axis of the microtubule and plotted against time. (b) Standard deviation of the raw data in (a), calculated using a 50-point (1-ms) moving window. (c) Wavelet-filtered s.d. data. The beginning and end of each event were determined by thresholding the wavelet-filtered data (dashed line; from Ref. [16]).

class of tetrameric kinesins, the Kinesin-5s, in particular the Eg5 motor of Xenopus laevis [9,10,18]. Eg5 has two pairs of motor-domains at each end of an extended tetrameric coiled-coil. Its cellular function is to aid in the morphogenesis of bipolar mitotic spindle during cell division. We could show that Eg5 dynamically crosslinks microtubules and thus provides the forces necessary to slide the spindle poles apart. We used an *in vitro* assay with purified Eg5 and fluorescently labeled microtubules. We bound bundles of microtubules (axonemes) to a glass coverslip, added motors and more microtubules. We found that single microtubules readily bound to and aligned with axonemes in the presence of Eg5 [18]. Approximately half of the microtubules were immotile or moved very slowly (< 10 nm/s) whereas the rest moved along the axonemes with an average velocity of 40 nm/s. With polarity marked microtubules as both tracks and substrate we could prove that only anti-parallel microtubules displayed relative motility. The relative movement of microtubules that were not aligned parallel showed that the motors could move simultaneously with respect to both linked microtubules (Fig. 12, Ref. [18]).

While these assays clearly indicated the capability of Eg5-kinesin to cross-linking and driving anti-parallel microtubules it was not clear whether the motility was driven by single Eg5-tetramers or by functional aggregates or patches of Eg5 at the cross-link

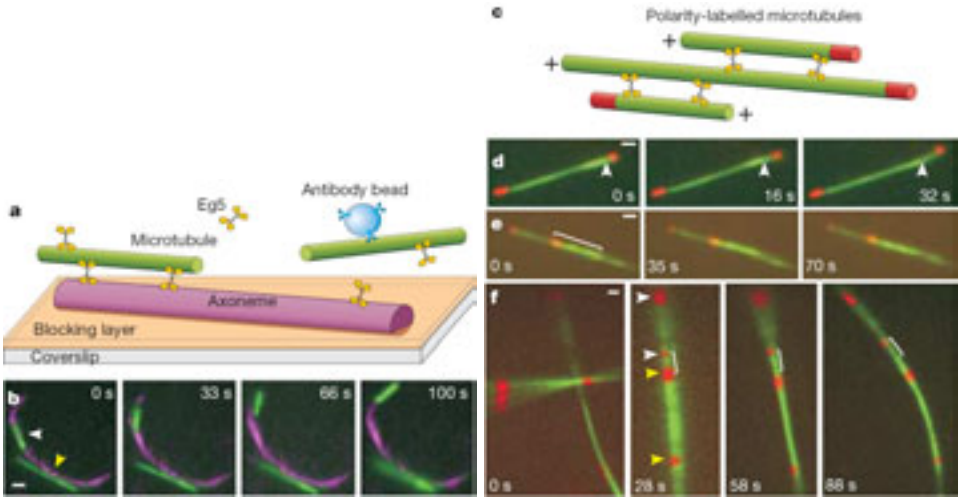


Figure 12. Eg5 can slide microtubules apart. (a) Sketch of the in vitro assay with microtubules (green) attached via Eg5 motors (yellow) to surface-immobilized axonemes (magenta). The coverslip surface is blocked using a polymer brush. Beads (1-mm diameter, blue) coated with anti-tubulin antibodies were used in some experiments for manipulation with optical tweezers. (b) Time-lapse images of both a sliding (white arrow, 40 nm/s) and a static (yellow arrow) fluorescent microtubule on a darkfield-detected axoneme. (c) Sketch of the *in vitro* assay with polarity-marked microtubules. (d) Antiparallel microtubules sliding apart. The arrow marks the plus-end of the long microtubule, relative to which the short one moved at 35 nm/s. (e) Two parallel microtubules (one marked with a white line) that were crosslinked and remained static. (f) Time course of sliding within a bundle of microtubules. Two bundles first joined and aligned (left panel). Seeds marked with arrows of the same colour remained stationary relative to each other, but moved at 36 nm/s relative to those marked with a different colour. Scale bar: 1 mm. (from Ref. [18]).

point. We therefore turned to single-molecule fluorescence experiments with GFP-labeled Eg5. Experiments showed that single Eg5-tetramers could interact for several tens of seconds with a microtubule, but their motility appeared irregular, a mixture of directed motion with diffusive intervals (Fig. 13). The buffer conditions influenced the prevalence of directed motion, low salt made the motors more directional. Under these conditions, the drug Monastrol again turned the directionality of the motor off. Interestingly, the motor was turned on even at higher salt concentration when it cross-linked two microtubules. Since these higher salt concentrations were close to physiological conditions, it is likely that Eg5 is regulated in such a way that it only is turned on when it is primed to do useful work, i. e. is bound between two microtubules. The basis of this activation remains unclear, but it is likely that the tail of one dimer interacts with the neck linker of the opposing dimer to effect this regulation [10].

A further finding that will help to understand the regulation of Eg5 was that an artificial chimera of Kinesin-1 and Kinesin-5, constructed from the motor domain and

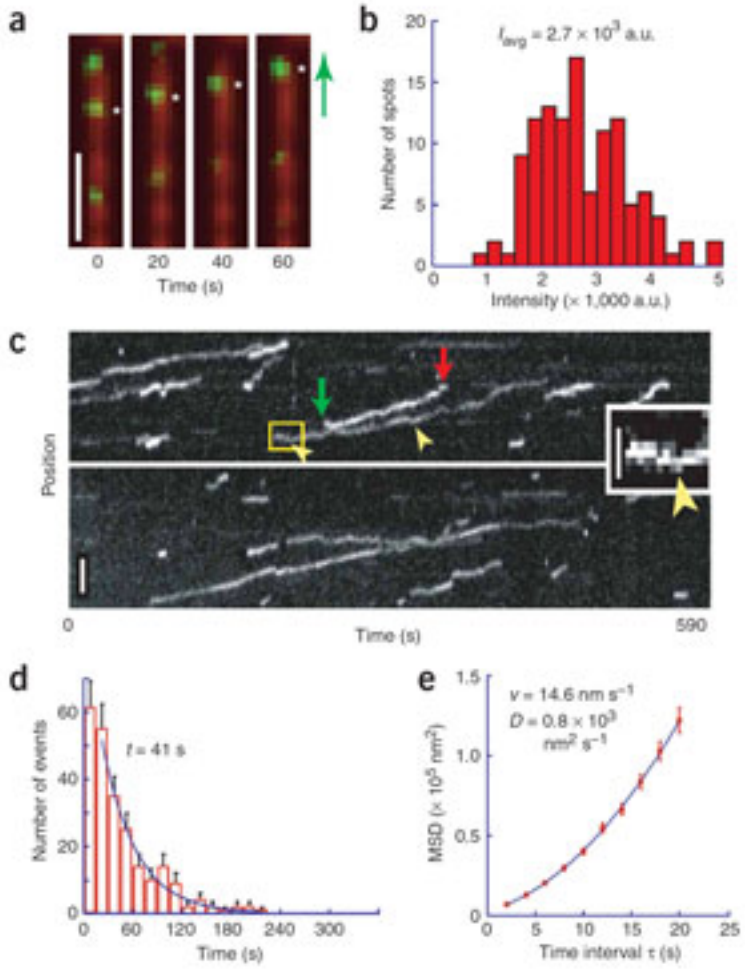


Figure 13. Full-length tetrameric Eg5 is a processive kinesin. (a) Frames from time-lapse recordings showing Eg5-GFP (green) moving along a microtubule (red). An asterisk (*) highlights one Eg5-GFP tetramer. The direction of motor movements is indicated by the green arrow. Bar, 2 mm. (b) Histogram of initial intensities of moving Eg5-GFP spots (n = 116). Mean intensity (I_{avg}) is indicated. (c) Kymographs depicting the motion of Eg5-GFP along microtubules in the presence of ATP (2 mM). The starting and ending points of a run are indicated by the green and the red arrows, respectively. Two examples of irregularities in the directional motility (that is, reversal in direction) are marked with yellow arrowheads. Bar, 2 mm. Inset: 3× magnifications of the framed area. Bar, 1 mm. (d) Histogram for the durations of Eg5-GFP-microtubule interactions of individual runs fitted by a single exponential. Average duration (t) is indicated (n = 239). (e) MSD calculated from Eg5-GFP motility recordings.The solid curve is a fit to MSD = $1/4v^2t^2 + 2Dt +$ offset. Values of v and D are indicated ($n = 80$). (From Ref. [10]).

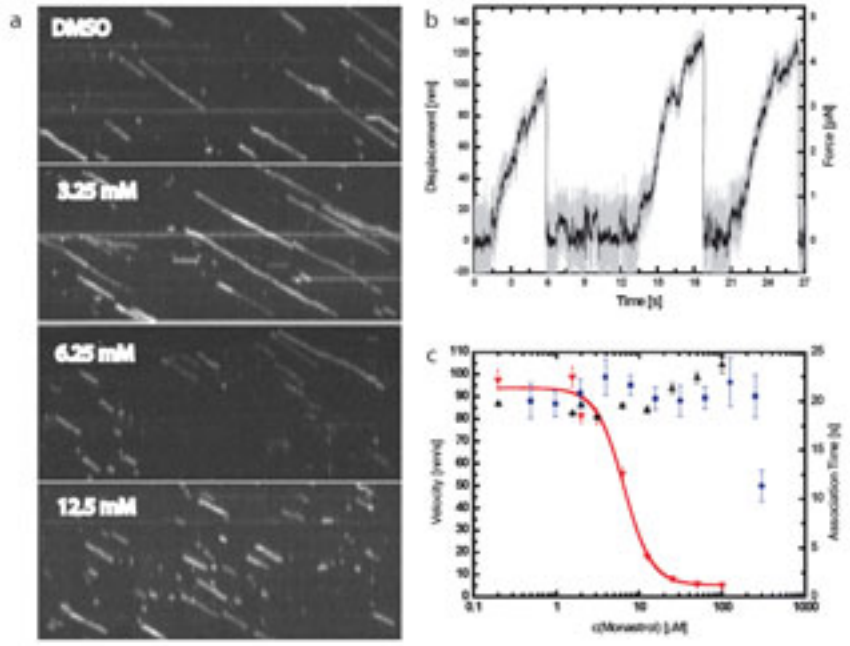


Figure 14. Motility and inhibition of single and multiple Eg5Kin motors. (a) Displacement and force produced by single truncated, C-terminally GFP-tagged Eg5Kin (Eg5Kin-GFP) motors. Eg5Kin-GFP motors were sparsely covered on silica glass spheres and presented to a microtubule using a single beam optical trap ($g = 0.035\,\mathrm{pN/nm}$). Eg5Kin-GFP moved the bead processively out of the trap center producing an average force of $4.6 \pm 0.1\,\mathrm{pN}$. Detachment occurs without observable stalling plateaus. (b) Kymograph of single Eg5Kin-GFP, moving for micrometer-long distances along a TMR-labeled microtubule with an average speed of 95 nm/s. Incremental, two-step bleaching was observed at points indicated by arrows, quantitatively confirming the dimeric status of Eg5Kin (x-axis=325 s, y-axis=11.32 μm). (c) Kymographs of single, GFP-tagged Eg5Kin motors moving along microtubules at increasing Monastrol concentrations (x-axis=200 s, y-axis=10.53 μm). (d) Graphical summary of motility data at increasing Monastrol concentrations: Eg5Kin-GFP single molecule association time (red triangle down, IC50 = 6.5 μM), Eg5Kin-GFP single molecule speeds (black triangles up), Eg5Kin multi motor surface gliding speeds (blue filled squares). (From Lakämper et al., manuscript in preparation).

neck-linker of Eg5 attached to the stalk of Kinesin 1 moved in a highly processive fashion along the microtubule for much longer distances than native Eg5. Monastrol reduced the run-length, but neither the speed nor the binding frequency of single dimeric chimeras (Fig. 14).

Fluorescence experiments are limited in two important ways. First, the number of photons emitted by a single dye molecule on an individual motor protein is so small that fast and accurate position detection is not possible. Second, one can neither

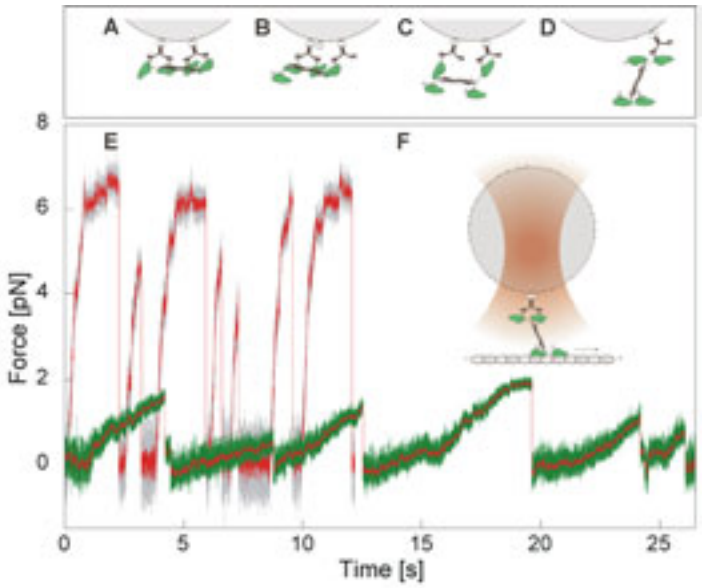


Figure 15. Motor attachment, experimental setup and bead traces. (a)–(d), (f) Possible Eg5-motor attachments to beads (silica, 0.5 μm diameter) via the genetically encoded N-terminal His-tag of Eg5. (a) All four motor domains are bound, preventing motility. (b) One motor domain is unbound, likely allowing only non-processive motility. (c) Two motor domains are free, one at each end, likely allowing only non-processive motility. (d) One motor domain is bound, leaving a dimeric motor end free to interact with the microtubule. (e) Traces of bead motility generated by individual Eg5 (green) and Kinesin-1 (grey) motors. The averaged (15-point) and median-filtered (0.3 s (Eg5) and 0.05 s (Kinesin-1) sliding windows; rank 10) signal is overlaid in red over both traces. The trap stiffnesses were 0.03 pN/nm (Kinesin-1) and 0.013 pN/nm (Eg5). (f) One dimer is bound, one dimer is free; sketch of a silica-sphere with motor held in the laser-trap, such that it interacts with a surface-attached microtubule track (from Ref. [9]).

exert any force on the motor, nor measure the force exerted by the motor. Both can be done with optical trapping assays where single motors are attached to optically trapped micron-sized beads. Motors of the Kinesin-1 class have been shown by such assays to move in 8 nm steps and exert maximal forces of about 7 pN. Figure 15 shows the comparison between Kinesin-1 and Eg5 motility [9]. Single Kinesin-1 dimers move for tens to hundreds of steps and eventually stall at about 6 pN load from the trap.

Eg5, in contrast, shows quite different motility behaviour: the motors moved less regularly than Kinesin-1, and they released at a load of typically below 2 pN. Nevertheless it was possible to discern 8 nm steps in the motion, confirming that the motors move in a fundamentally similar manner to Kinesin-1 motors. Our findings suggest that full-length Eg5-tetramers might employ a so far undescribed mechanism to limit force-production of individual motors – a sort of slip-clutch-mechanism – which might have a role in regulating spindle dynamics [9].

3.4 Fluid dynamics, polymer networks, colloids and model systems for cells studied by microrheology

The machinery that drives essential functions of cells such as locomotion and division is based on an elastic network of interconnected semiflexible protein filaments, collectively referred to as the cytoskeleton. A major component of the cytoskeleton is the actin cortex, a dense meshwork of cross-linked actin filaments beneath the plasma membrane that is controlled by a host of accessory proteins. The physical construction of the cytoskeleton with its complex hierarchy of structural length scales enables the cell to produce large changes in physical properties by small chemical interventions, such as length- or crosslink-control or regulated attachments to other structures in the cell. The unique sensitivity of cytoskeletal networks stems in large part from the semiflexible character of its constituents, i. e., the fact that their thermal persistence length l_p (17 μm for filamentous-actin (F-actin)) is orders of magnitude larger than molecular scales (7 nm filament diameter of actin). The mechanical and dynamical (rheological) properties of semiflexible polymers have been the focus of intense research in recent years. Apart from their biological role, these networks have proven to be unique polymeric materials in their own right. In contrast to flexible polymer networks, the shear modulus of a semiflexible polymer network can be varied over many orders of magnitude by small changes in cross-linking, and exhibits strong nonlinearities. The dynamics of semiflexible solutions and gels have proven to be much richer than those of flexible polymers. Even for single filaments there are multiple distinct modes of relaxation that are qualitatively distinct from those of conventional polymers. It has proven challenging, however, to quantitatively probe those dynamic regimes experimentally, because of the extensive bandwidth required [11].

Microrheology based on optical traps and interferometric detection of particle motions can meet those challenges. Having a bandwidth of 6 orders of magnitude in frequency from 0.1 Hz to 100 kHz, however, provides other interesting options. It makes it also possible to study general issues of fluid dynamics. A fundamental problem in hydrodynamics is the response of a liquid to the motion of a small embedded particle. At sufficiently long times, the well known Stokes velocity field, which decreases as $1/r$ away from the particle, will describe this fluid response. For an initial disturbance due to a local force in the liquid, however, only a small region of the liquid can be set in motion due to the inertia of the liquid. If the liquid is incompressible, backflow occurs that is characterized by a vortex ring surrounding the point disturbance. Since vorticity diffuses within the (linearized) Navier-Stokes equation, propagation of shear in the fluid drives the expansion of this vortex ring as a function of time t as $\sqrt{t}$. The $1/r$ Stokes flow is established only in the wake of this vortex. While this basic picture has been known theoretically for simple liquids since Oseen [12], and has been observed in simulations since the 1960's [19], this vortex flow pattern has not been observed directly in experiment. In a recent project we have used the correlations in thermal fluctuations of small probe particles to resolve this vortex flow field on the micrometer scale along with its diffusive propagation. We found good agreement between measured flow patterns and theoretical calculations for simple viscous fluid. Furthermore, we could demonstrate similar vortex-like flow in viscoelastic media. In the viscoelastic case, interestingly, vorticity spreads super-diffusively.

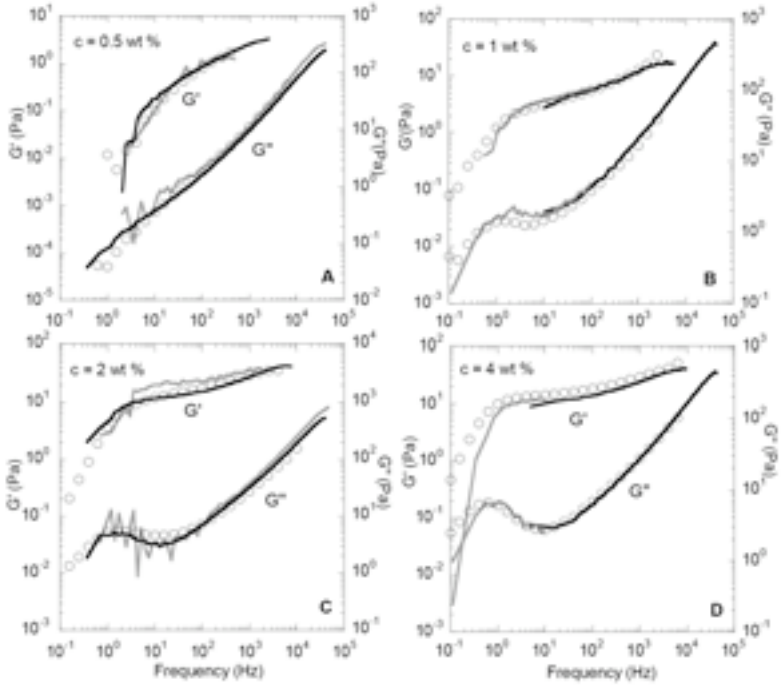


Figure 16. Loss modulus (right axes) and elastic modulus (left axes) for four surfactant concentrations 0.5 (A), 1 (B), 2 (C), and (D) 4 wt % are plotted as a function of frequency. Curves are in (A) and (C): macrorheology (circles), 1PMR (black lines), and 2PMR (gray lines) in 0.5 and 2 wt %; in (B) and (D): 1PMR with 20 kHz sampling rate (gray line) and 1PMR with 195 kHz sampling rate (black lines). All microrheology data were logarithmically binned with the factor of 1.2 relating the widths of successive bins. (from Ref. [21]).

To characterize the new microrheology methods we have been developing in their performance on polymer samples we performed a rigorous comparison between established conventional rheology and microrheology on a stable, well known system namely entangled solutions of wormlike micelles which behave like a simple Maxwell fluid at low frequencies. To generate enough of an overlap in bandwidth between macro- and microrheology we used a specialized design based on piezoelectric actuators for the macrorheology. Wormlike micelles are cylindrical assemblies of amphiphilic molecules that form spontaneously in aqueous solutions at particular concentrations and temperature conditions. We have used cetylpyridinium chloride (CPyCl) as the surfactant and sodium salicylate (NaSal) as a strongly binding counterion. The wormlike micelles formed in this system have a diameter of 2 to 3 nm, contour lengths of 100 nm to 1 μm, and a persistence length of order 10 nm. At the concentrations we used, the mesh size varied from about 0 to 10 nm [20,21].

We have compared one-particle and two-particle microrheology with macrorheology. With all three techniques we have obtained frequency dependent complex shear moduli over large and overlapping frequency ranges (Fig. 16). Excellent agreement

of the results from all three techniques was observed. This was in principle not unexpected given that characteristic length scales of the solution, such as persistence length and mesh size, were significantly smaller than the probe particle size. Our results provided a much needed quantitative verification of microrheology on a simple model system [20,21].

With the approach validated, results on more biologically relevant systems could be understood. We initially examined in vitro reconstituted networks of entangled filamentous actin by one-particle passive microrheology. A main result was the high-frequency scaling behaviour of the shear modulus $G^*(\omega) \sim \omega^{3/4}$ with a power law exponent that is characteristic for semiflexible polymer networks [11]. Probing with single particles is, however, likely to misreport the actual bulk shear viscosity of the embedding medium if there are characteristic length scales of the medium that are comparable to the probe size. This is the case for actin networks, and an effect that tends to make the measured shear modulus lower than the true bulk value is steric depletion of the network around the probe particle. This problem can be avoided by evaluating the correlated fluctuations of a particle pair in two-particle passive microrheology. With this technique we again analyzed actin solutions and obtained quantitative agreement with theoretical predictions of the shear modulus in both amplitude and frequency dependence (Fig. 17, Ref. [22]).

In systems that are in thermodynamic equilibrium, active and passive microrheology should give the same results. In systems out of equilibrium, though, the combination of active and passive microrheology can be employed to characterize non-thermal fluctuations. Developing a statistical mechanical description of non-equilibrium systems such as glasses still remains an important challenge in physics [24]. One of the most interesting recent developments along these lines is the proposal to generalize the fluctuation dissipation theorem (FDT) to non-equilibrium situations. The FDT relates the response of a system to a weak external perturbation to the relaxation of the spontaneous fluctuations in equilibrium. The response function is proportional to the power spectral density of thermal fluctuations, with a prefactor given by the temperature. This suggests a generalization for systems out of equilibrium, in which the (non-equilibrium) fluctuations are related to the response via a time-scale-dependent effective temperature. While this has been studied extensively theoretically, the experimental support for a meaningful effective temperature is unclear. There have been few experiments and the usefulness of the extension of the FDT to non-equilibrium situations is still a matter of controversy. We have used a combination of active and passive (fluctuation-based) microrheology techniques that provide a way to directly test the applicability of the FDT. We have examined the validity of the FDT in a colloidal glass, the synthetic clay Laponite. For this system conflicting results had been reported previously, that may in part have been due to the use of a limited experimental window in both frequency and aging time. We have performed measurements over a wide range of frequencies and aging times. Contrary to previous reports, we find no violation of the FDT and thus no support for an effective temperature different from the bath temperature [24].

While the aging colloidal glasses are changing very slowly and are therefore not very strongly non-equilibrium, a living cell shows much stronger signatures of energy dissipation, i. e. non-equilibrium dynamics. Many cellular functions such as cell

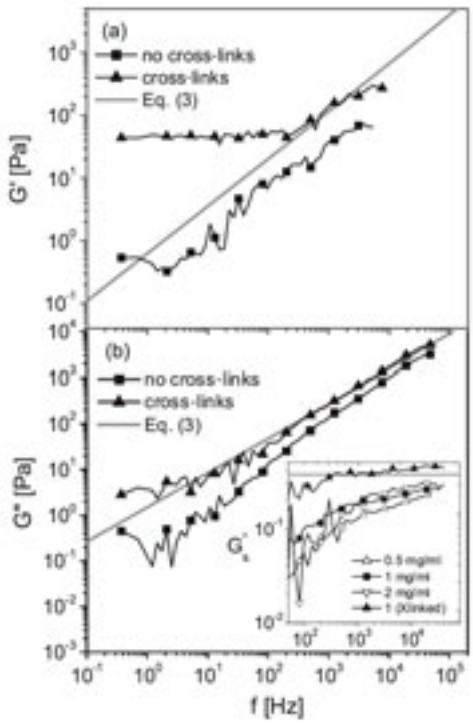


Figure 17. (a) Storage modulus $G'(\omega)$ and (b) (absolute) loss modulus $G''(\omega)$ of 1 mg/ml solutions of F-actin filaments without (squares) and with (triangles) cross-linking plotted against frequency $f = \omega/2\pi$. Solid lines: theoretical modelling. (From Ref. [22]).

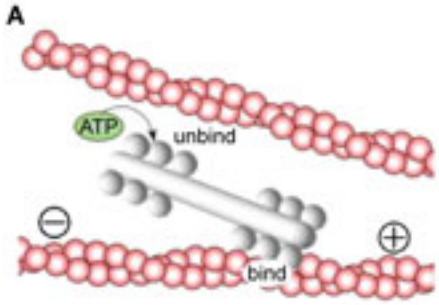


Figure 18. Schematic drawing of a bipolar myosin filament interacting with two actin filaments. Polarity of actin is indicated by the +/− signs (myosin moves toward the plus end; from Ref. [23]).

locomotion or cell division involve movement and rearrangements of the cell structures that occur on the scale of seconds to minutes. The cytoskeleton is a network of semiflexible linear protein polymers (actin filaments, microtubules, and intermediate filaments) that is responsible for most of the mechanical functions of cells. It differs from common polymer materials in both the complexity of composition and the fact that the system is not in thermodynamic equilibrium. Chemical non-equilibrium drives mechanoenzymes (motor proteins) that are the force generators in cells. The cytoskeleton is thus an active material that can adapt its mechanics and perform mechanical tasks such as cell locomotion or cell division. We have explored in a recent project how non-equilibrium motor activity controls the mechanical properties of a simple three-component *in vitro* model cytoskeletal network consisting of a cross-linked actin network with embedded force-generating myosin II motors which are the skeletal muscle motors [23].

We formed myosin "minifilaments" (Fig. 18) that can link different actin filaments and move these filaments relative to each other in the presence of ATP. In the absence of ATP, these motor complexes statically cross-link F-actin and generate bundles that can be seen in a light microscope (data not shown here). In the presence of ATP, minifilaments generate contractile forces that can result in actin aggregation and phase separation, a phenomenon known as superprecipitation. To stabilize the networks and delay the onset of superprecipitation, we used F-actin cross-linked

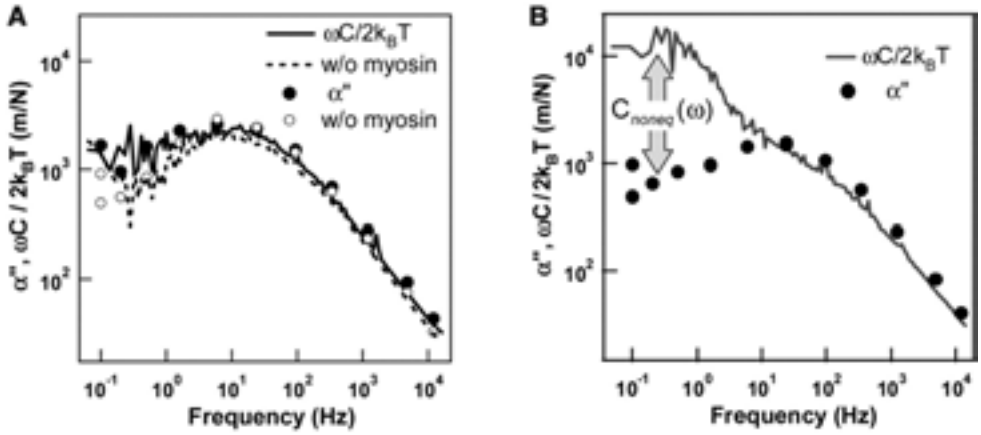


Figure 19. Mechanical response of cross-linked nonactive and active gels (actin and myosin concentrations as in Fig. 1). (A) The imaginary part of the response function α'' measured by AMR (circles) and the normalized power spectrum $\omega C(\omega)/2k_BT$ measured by PMR (lines). Open circles and the dashed line denote cross-linked actin without myosin; solid circles and the solid line denote networks with myosin 2.5 hours after sample preparation. For up to 5 hours, α'' and $\omega C(\omega)/2k_BT$ with and without myosin show good agreement, indicating that myosin activity did not yet produce observable non-equilibrium fluctuations. (B) The same as (A) but 6.8 hours after sample preparation (with myosin). Below 10 Hz, non-equilibrium fluctuations are observable as an enhancement of $\omega C(\omega)/2k_BT$ relative to α'' (from Ref. [23]).

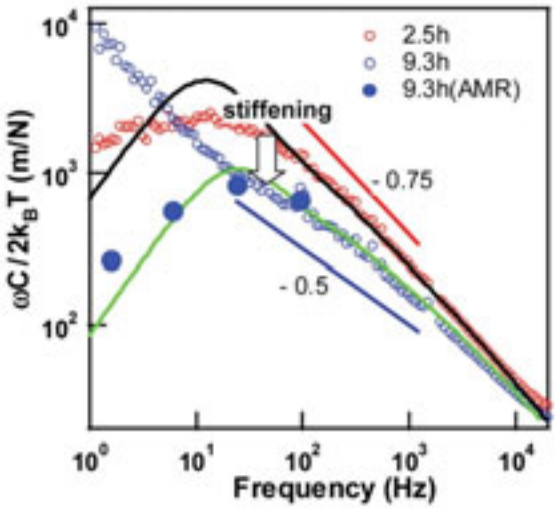


Figure 20. Effect of filament tension on the response of the active networks (actin and myosin concentrations as in Fig. 19. Spectra $\omega C(\omega)/2k_BT$ measured with PMR at 2.5 hours (open red circles) and 9.3 hours (open blue circles) and α'' measured with AMR at 9.3 hours (solid blue circles) after sample preparation (initial [ATP] = 3.5 mM). In the presence of non-equilibrium activity, the response function is reduced, indicating a stiffer sample, which can be fully accounted for by prestress/tension of filaments. Theoretical predictions are shown for a network with filament tension of 0.1 pN, cross-link distance $l_c = 2.6\,\mu$m (green curve), and no tension with the same l_c (black curve). Independently known parameters: friction coefficient $z = 0.00377$ Pa/s, persistence length $l_p = 17 \times 10^{-6}$ m, probe radius $a = 2.5$ mm. The system strongly violates the FD theorem and that it does so because of the contractility of the acto-myosin system [23].

by biotin and neutravidin. We then measured the mechanical properties of these networks by active microrheology and found agreement with previous data from cross-linked actin networks (Fig. 19, Ref. [23]).

Passive microrheology gave results that agreed completely. With ATP-energized myosin motors, however, the active processes created additional fluctuations and thus violated the FD theorem [23]. The motor-generated tensions in the network also dramatically boosted the shear modulus, by up to a factor of 100 (Fig. 20).

Thus, actin, myosin, and cross-links are sufficient to capture essential and general features of contractility and mechanical adaptation in cytoskeletal networks. These observations suggest mechanisms by which cells could rapidly modulate their stiffness by flexing their internal "muscles" without changing the density, polymerization, or bundling state of F-actin. Cells can actively adapt their elasticity to the mechanics of the extracellular matrix or to an externally applied force, and motors could be the cause for that. From a materials perspective, this in vitro model system exhibits an active state of matter that adjusts its own mechanical stiffness via internal forces. This work can be a starting point for exploring both model systems and cells in

quantitative detail, with the aim of uncovering the physical principles underlying the active regulation of the complex mechanical functions of cells.

References

[1] I. A. Schaap, P. J. de Pablo, and C. F. Schmidt, 'Resolving the molecular structure of microtubules under physiological conditions with scanning force microscopy', Eur. Biophys. J. **33**, 462 (2004).

[2] P. J. de Pablo, I. A. Schaap, F. C. MacKintosh, and C. F. Schmidt, 'Deformation and collapse of microtubules on the nanometer scale', Phys. Rev. Lett. **91**, 098101 (2003).

[3] I. A. Schaap, B. Hoffmann, C. Carrasco, R. Merkel, and C. F. Schmidt, 'Tau protein binding forms a 1 nm thick layer along protofilaments without affecting the radial elasticity of microtubules', J. Struct. Biol. **158**, 282 (2007).

[4] I. A. Schaap, C. Carrasco, P. J. de Pablo, F. C. MacKintosh, and C. F. Schmidt, 'Elastic response, buckling, and instability of microtubules under radial indentation', Biophys. J. **91**, 1521 (2006).

[5] J. P. Michel, I. L. Ivanovska, M. M. Gibbons, W. S. Klug, C. M. Knobler, G. J. Wuite, and C. F. Schmidt, 'Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength', Proc. Natl. Acad. Sci. USA **103**, 6184 (2006).

[6] W. S. Klug, R. F. Bruinsma, J. P. Michel, C. M. Knobler, I. L. Ivanovska, C. F. Schmidt, and G. J. Wuite, 'Failure of viral shells', Phys. Rev. Lett. **97**, 228101 (2006).

[7] I. L. Ivanovska, P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. Wuite, 'Bacteriophage capsids: tough nanoshells with complex elastic properties', Proc. Natl. Acad. Sci. USA **101**, 7600 (2004).

[8] R. P. Goodman, I. A. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, and A. J. Turberfield, 'Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication', Science **310**, 1661 (2005).

[9] M. J. Korneev, S. Lakämper, and C. F. Schmidt, 'Load-dependent release limits the processive stepping of the tetrameric Eg5 motor', Eur. Biophys. J. **36**, 675 (2007).

[10] B. H. Kwok, L. C. Kapitein, J. H. Kim, E. J. Peterman, C. F. Schmidt, and T. M. Kapoor, 'Allosteric inhibition of kinesin-5 modulates its processive directional motility', Nat. Chem. Biol. **2**, 480 (2006).

[11] F. C. MacKintosh and C. F. Schmidt, 'Microrheology', Opin. Coll. Interf. Sci. **4**, 300 (1999).

[12] C. W. Oseen, *Neuere Methoden und Ergebnisse in der Hydrodynamik* (Akad. Verl.-Ges., Leipzig, 1927).

[13] S. Lakämper and E. Meyhofer, 'Back on track – on the role of the microtubule for kinesin motility and cellular function', J. Muscle Res. Cell Motil. **27**, 161 (2006).

[14] S. Lakämper and E. Meyhofer, 'The E-hook of tubulin interacts with kinesin's head to increase processivity and speed', Biophys. J. **89**, 3223 (2005).

[15] S. Lakämper, A. Kallipolitou, G. Woehlke, M. Schliwa, and E. Meyhofer, 'Single fungal kinesin motor molecules move processively along microtubules', Biophys. J. **84**, 1833 (2003).

[16] M. J. de Castro, R. M. Fondecave, L. A. Clarke, C. F. Schmidt, and R. J. Stewart, 'Working strokes by single molecules of the kinesin-related microtubule motor ncd', Nat. Cell Biol. **2**, 724 (2000).

[17] M. W. Allersma, F. Gittes, M. J. de Castro, R. J. Stewart, and C. F. Schmidt, 'Two-

dimensional tracking of ncd motility by back focal plane interferometry', Biophys. J. **74**, 1074 (1998).
[18] L. C. Kapitein, E. J. Peterman, B. H. Kwok, J. H. Kim, T. M. Kapoor, and C. F. Schmidt, 'The bipolar mitotic kinesin Eg5 moves on both microtubules that it crosslinks', Nature **435**, 114 (2005).
[19] B. J. Alder and T. E. Wainwright, 'Velocity autocorrelations for hard spheres', Phys. Rev. Lett. **18**, 988 (1967).
[20] M. Buchanan, M. Atakhorrami, J. F. Palierne, F. C. MacKintosh, and C. F. Schmidt, 'High-frequency microrheology of wormlike micelles', Phys. Rev. E **72**, 011504 (2005).
[21] M. Buchanan, M. Atakhorrami, J. F. Palierne, and C. F. Schmidt, 'Comparing macrorheology and one- and two-point microrheology in wormlike micelle solutions', Macromolecules **38**, 8840 (2005).
[22] G. H. Koenderink, M. Atakhorrami, F. C. MacKintosh, and C. F. Schmidt, 'High-frequency stress relaxation in semiflexible polymer solutions and networks', Phys. Rev. Lett. **96**, 138307 (2006).
[23] D. Mizuno, C. Tardin, C. F. Schmidt, and F. C. MacKintosh, 'Nonequilibrium mechanics of active cytoskeletal networks', Science **315**, 370 (2007).
[24] S. Jabbari-Farouji, D. Mizuno, M. Atakhorrami, F. C. MacKintosh, C. F. Schmidt, E. Eiser, G. H. Wegdam, and D. Bonn, 'Fluctuation-dissipation theorem in an aging colloidal glass', Phys. Rev. Lett. **98**, 108302 (2007).

Index

A broad variety of research topics emerged during the past sixty years from the institute's global theme „oscillations and waves". Some of these topics are addressed in this book in which topical review articles by former and present members of the institute are collected. The subjects covered vary from speech and hearing research to flow control and active control systems, from bubble oscillations to cavitation structures, from ordering phenomena in liquids and one-dimensional solids to complex dynamics of chaotic nonlinear systems, from laser speckle metrology to ring laser gyroscopes, from biophysics to medical applications in ophthalmology as well as extracorporeal shock wave lithotripsy.

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

ISBN-13: 978-3-938616-96-3

Universitätsverlag Göttingen