PHYSICAL ACOUSTICS SUMMER SCHOOL

30 Sep 2001

PASS 2000

ASILOMAR CONFERENCE CENTER

VOLUME I

TRANSCRIPTS

2000 PHYSICAL ACOUSTICS SUMMER SCHOOL (PASS 00)

VOLUME I: TRANSCRIPTS

This work relates to Department of Navy Grant 14-98-1-0044 issued by the Office of Naval Research.

Copies of this three-volume proceedings can be obtained by contacting: Libby Furr, NCPA, University of Mississippi, University MS 38677; voice: 662-915-5808; fax: 662-915-7494; e-mail: libby@olemiss.edu

TABLE OF CONTENTS

2000 PHYSICAL ACOUSTICS SUMMER SCHOOL

FORWARD

These are the Proceedings of the 2000 Physical Acoustics Summer School (PASS 00). The lectures were recorded and most of the verbatim transcripts were subsequently edited by the authors for publication here.

Sponsored by the Office of Naval Research (ONR) and organized in cooperation with the Acoustical Society of America (ASA) and the National Center for Physical Acoustics (NCPA), PASS 00 was held June 16-23, 2000 at the Asilomar Conference Center in Pacific Grove, California, the site of the first and subsequent Summer Schools, PASS 92, PASS 94, PASS 96 and PASS 98. Participation in each of the Summer Schools was limited to a total of 50 that included students, lecturers, and discussion leaders.

The purpose of these Summer Schools is to bring graduate students, distinguished lecturers, and discussion leaders together to discuss a wide variety of subjects in physical acoustics. This gives the students the opportunity to meet experts and talk about topics most students ordinarily wouldn't encounter at their own colleges and universities. The focus was on graduate students and academic participants. Approximately half of the participants have been advanced graduate students in physical acoustics.

The Summer Schools have their beginning in an ONR Principal Investigators meeting in 1988 where it was decided that the best investment of this kind would be in a Summer School focused on graduate students. In 1990, an informal Summer School for students of ONR Principal Investigators and some invited guests was held as part of a Principal Investigator's meeting at the Naval Postgraduate School, Monterey, California, and this controlled experiment set the pattern and influenced the site selection for the subsequent PASS 92, PASS 94, PASS 96, PASS 98 and PASS 00. We hope that the Summer Schools will continue as biennial events with the high standards and wonderful success we have thus far enjoyed. This has happened and will happen again because everyone involved does their best, and for this I say, "Thank You!"

LOGAN E. HARGROVE – ONR 331

2000 PHYSICAL ACOUSTICS SUMMER SCHOOL
LIST OF PARTICIPANTS

-------LECTURERS-------

ANTHONY ARMSTRONG ATCHLEY
PENNSYLVANIA STATE UNIVERSITY
GRADUATE PROGRAM IN ACOUSTICS
P.O. BOX 30
STATE COLLEGE, PA 16804
PHONE: 814-865-6364
EMAIL: AATCHLEY@PSU.EDU

LAWRENCE A. CRUM
UNIVERSITY OF WASHINGTON
APPLIED PHYSICS LABORATORY
1013 NE 40TH STREET
SEATTLE, WA 98105
PHONE: 206-685-8622
EMAIL: LAC@APL.WASHINGTON.EDU

BRUCE C. DENARDO
NAVAL POSGRADUATE SCHOOL
PHYSICS DEPARTMENT
CODE PH/DE
MONTEREY, CA 93943
PHONE: 831-656-2952
EMAIL: DENARDO@PHYSICS.NPS.NAVY.MIL

THOMAS B. GABRIELSON
PENNSYLVANIA STATE UNIVERSITY
APPLIED RESEARCH LABORATORY
P.O. BOX 30
STATE COLLEGE, PA 16804
PHONE: 814-865-1370
EMAIL: TBG3@PSU.EDU

STEVEN LURIE GARRETT
PENNSYLVANIA STATE UNIVERSITY
GRADUATE PROGRAM IN ACOUSTICS
P.O. BOX 30
STATE COLLEGE, PA 16804-0030
PHONE: 814-863-6373
EMAIL: GARRETT@SABINE.ACS.PSU.EDU

KENNETH E. GILBERT
UNIVERSITY OF MISSISSIPPI
NATIONAL CENTER FOR PHYSICAL ACOUSTICS
COLISEUM DRIVE
UNIVERSITY, MS 38677
PHONE: 662-915-5634
EMAIL: KGILBERT@OLEMISS.EDU

ROBERT MITCHEL KEOLIAN
PENNSYLVANIA STATE UNIVERSITY
GRADUATE PROGRAM IN ACOUSTICS
P.O. BOX 30
STATE COLLEGE, PA 16804
PHONE: 814-865-1365
EMAIL: BONZO@SABINE.ACS.PSU.EDU

TOM JOHN MATULA
UNIVERSITY OF WASHINGTON
APPLIED PHYSICS LABORATORY
1013 NE 40TH STREET
SEATTLE, WA 98105
PHONE: 206-685-7654
EMAIL: MATULA@APL.WASHINGTON.EDU

ALBERT MIGLIORI
LOS ALAMOS NATIONAL LABORATORY
MS E561
LOS ALAMOS, NM 87545
PHONE: 505-667-2515
EMAIL: MIGLIORI@LANL.GOV

JAMES M. SABATIER
UNIVERSITY OF MISSISSIPPI
NATIONAL CENTER FOR PHYSICAL ACOUSTICS
COLISEUM DRIVE
UNIVERSITY, MS 38677
PHOEN: 662-915-5404
EMAIL: SABATIER@OLEMISS.EDU

----Discussion Leaders----

William Patrick Arnott
Desert Research Institute
Atmospheric Sciences Center
2215 Raggio Parkway
Reno, NV 89215-1095
phone: 775-674-7023
email: pat@dri.edu

Henry Ellis Bass
University of Mississippi
National Center for Physical Acoustics
1 Coliseum Drive
University, MS 38677
phone: 662-915-5840
email: pabass@olemiss.edu

James Patrick Chambers
University of Mississippi
National Center for Physical Acoustics
Coliseum Drive
University, MS 38677
phone: 662-915-5100
email: chambers@olemiss.edu

R. Daniel Costley
MILTEC Inc.
National Center for Physical Acoustics
University of Mississippi
University, MS 38677
voice: 662-915-5692
email: dcostley@mil-tec.com

Mark F. Hamilton
University of Texas @ Austin
Department of Mechnical Engineering
Campus Mail Code C220
Austin, TX 78712
phone: 512-471-3055
email: hamilton@mail.utexas.edu

Logan E. Hargrove
Office of Naval Research ONR 331
800 North Quincy Street
Arlington, VA 22217-5660
phone: 703-696-4221
email: hargrol@onr.navy.mil

Veerle M. Keppens
University of Mississippi
National Center for Physical Acoustics
Coliseum Drive
University, MS 38677
phone: 662-915-7428
email: vkeppens@olemiss.edu

Philip L. Marston
Washington State University
Department of Physics
Pullman, WA 99164-2814
phone: 509-335-5343
email: marston@wsu.edu

Julian Decatur Maynard
Pennsylvania State University
Department of Physics
104 Davey Lab
University Park, PA 16802
phone: 814-865-6353
email: maynard@phys.psu.edu

Victor W. Sparrow
Pennsylvania State University
Acoustics Program
157 Hammond Building
University Park, PA 16802
phone: 814-865-3162
email: sparrow@helmholtz.acs.psu.edu

Richard Stern
Pennsylvania State University
Applied Research Laboratory
P.O. Box 30
State College, PA 16804
phone: 814-865-6344
email: rstern@psu.edu

Stefan John Turneaure
Pennsylvania State University
Graduate Program in Acoustics
P.O. Box 30
State College, PA 16804
phone: 814-863-1899
email: sjt10@email.psu.edu

Doru Velea
National Center for Physical Acoustics
University of Mississippi
Coliseum Drive
University, MS 38677
phone: 662-915-5602
email: dvelea@olemiss.edu

Roger Martin Waxler
National Center for Physical Acoustics
University of Mississippi
Coliseum Drive
University, MS 38677
phone: 662-915-5259
email: rwax@olemiss.edu

Larry Alan Wilen
Ohio University
Department of Physics and Astronomy
167 Clippinger Laboratory
Athens, OH 45701
phone: 740-593-9610
email: wilen@helios.phy.ohiou.edu

--------Students--------

Antonios Apostolou
2418 S. Voss Rd. #L-205
Houston, TX 77057
phone: 281-285-7323
email: apostolou@sugar-land.oilfield.slb.com

Wen-Shiang Chen
University of Washington
5259 Mithun Place NE
Seattle, WA 98105
phone: 206-524-7004
email: wschen@u.washington.edu

Washington J. N. de Lima
Rua Luiz Oscar de Carvalho, 75
BL B-15, Apt. 34 Trindade
Florianopolis-SC, Brazil
88036-000
phone: 55 48 234-2973
email: washdelima@hotmail.com

Utkan Demirci
Stanford University
Crothers Hall 98B
Stanford, CA 94305
phone: 650-498-0835
email: utkan@leland.stanford.edu

Joseph Rhea Gladden
Pennsylvania State University
345 Oakley Drive
State College, PA 16803
phone: 814-865-7841
email: jgladden@psu.edu

Jennifer Eve Hightower
Colorado State University
905 West Laurel Street #114
Fort Collins, CO 80521
phone: 970-491-0872
email: jhightow@lamar.colostate.edu

Dustin Edward Kruse
University of California @ Davis
1959 Lake Boulevard
Apartment 157
Davis, CA 95616
phone: 530-752-7138
email: dekruse@ucdavis.edu

Tatsuro Nagano
Washington State University
Department of Physics
Pullman, WA 99164
phone: 509-335-7286
email: tnagano@wsunix.wsu.edu

Andi G. Petculescu
Ohio University
34 NorthMcKinley Avenue
Apartment 206
Athens, OH 45701
phone: 740-594-4308
email: apetcule@helios.phy.ohio.edu

GABRIELA PETCULESCU
OHIO UNIVERSITY
PHYSICS DEPARTMENT
CLIPPINGER LABS
ROOM 251
ATHENS, OH 45701
PHONE: 740-593-1757
EMAIL: PETCULES@HELIOS.PHY.OHIOU.EDU

SANDRA LOUISE POLIACHIK
UNIVERSITY OF WASHINGTON
APPLIED PHYSICS LABORATORY
1013 NE 40TH STREET
SEATTLE, WA 98105-6698
PHONE: 206-685-9648
EMAIL: POLIACHI@U.WASHINGTON.EDU

TYRONE MICHAEL PORTER
UNIVERSITY OF WASHINGTON
APPLIED PHYSICS LABORATORY
1013 NE 40TH STREET
SEATTLE, WA 98015
PHONE: 206-543-1323
EMAIL: TPORTER@APL.WASHINGTON.EDU

TIMOTHY G. SIMMONS
UNIVERSITY OF MISSISSIPPI
NATIONAL CENTER FOR PHYSICAL ACOUSTICS
COLISEUM DRIVE
UNIVERSITY, MS 38677
PHONE: 662-915-5843
EMAIL: TGSIMMON@OLEMISS.EDU

DAHLIA L. SOKOLOV
UNIVERSITY OF WASHINGTON
APPLIED PHYSICS LABORATORY
1013 NE 40TH STREET
SEATTLE, WA 98105
PHONE: 206-221-4778
EMAIL: DSOKOLOV@APL.WASHINGTON.EDU

MICHELLE ELAINE SWEARINGEN
506 SOUTHLAND CIRCLE DRIVE
TUSCOLA, IL 61953
PHONE: 217-253-9723
EMAIL: MICHELLE.SWEARINGEN@POBOX.COM

EDWARD J. TUCHOLSKI
NAVAL POSTGRADUATE SCHOOL
PHYSICS DEPT, RM 200 SPANAGLE HALL
833 DYER ROAD
MONTEREY, CA 93943
PHONE: 831-656-2020
EMAIL: EJTUCHOL@NPS.NAVY.MIL

BRIAN CRAIG TUTTLE
PENNSYLVANIA STATE UNIVERSITY
314 WEST NITTANY AVENUE
STATE COLLEGE, PA 16801
PHONE: 814-238-4223
EMAIL: BTUTTLE@SABINE.ACS.PSU.EDU

WEI WEI
WASHINGTON STATE UNIVERSITY
P.O. BOX 2512
PULLMAN, WA 99165-2512
PHONE: 509-335-7918
EMAIL: WEI_WEI@MAIL.WSU.EDU

JOHN R. WELCH
PO BOX 892
BUZZARD'S BAY, MA 02532
PHONE: 401-832-3889
EMAIL: JOHNWELCH_2000@YAHOO.COM

PRESTON SCOT WILSON
BOSTON UNIVERSITY
110 CUMMINGTON STREET
ROOM 101
BOSTON, MA 02215
PHONE: 617-353-4854
EMAIL: PSW@BU.EDU

XIANG YAN
MECHANICAL ENGINEERING
CAMPUS MAIL CODE: C2200
UNIVERSITY OF TEXAS
AUSTIN, TX 78712
PHONE: 512-471-5771
EMAIL: XIANGYAN@MAIL.UTEXAS.EDU

PASS 2000

PHYSICAL ACOUSTICS SUMMER SCHOOL

ASILOMAR CONFERENCE CENTER • PACIFIC GROVE CA • 16-23 JUNE 2000

TENTATIVE SCHEDULE

FRIDAY 16 JUNE	SATURDAY 17 JUNE	SUNDAY 18 JUNE	MONDAY 19 JUNE	TUESDAY 20 JUNE	WEDNESDAY 21 JUNE	THURSDAY 22 JUNE	FRIDAY 23 JUNE
	BREAKFAST 7:30 - 8:30	BREAKFAST 7:30 - 8:30	BREAKFAST 7:30 - 8:30	BREAKFAST 7:30 - 8:30	BREAKFAST 7:30 - 8:30	BREAKFAST 7:30 - 8:30	BREAKFAST 7:30 - 8:30
	A. A. ATCHLEY *Introduction to Physical Acoustics* 8:30 - 11:30	FREE MORNING	T. G. GABRIELSON *Noise and Sensors* 8:30 - 11:30	K. E. GILBERT *Atmospheric and Meteorological Acoustics* 8:30 - 11:30	R. M. KEOLIAN *Thermoacoustics* 8:30 - 11:30	L. A. CRUM *Medical Applications of Acoustics* 8:30 - 11:30	CHECK OUT BEFORE NOON
	LUNCH 12:00 - 1:00	LUNCH 12:00 - 1:00	LUNCH 12:00 - 1:00	LUNCH 12:00 - 1:00	LUNCH 12:00 - 1:00	LUNCH 12:00 - 1:00	LUNCH 12:00 - 1:00
REGISTRATION OPENS at 3:00	S. L. GARRETT *Nonlinear Acoustics* 1:30 - 4:30	B. C. DENARDO *Acoustics Demos and NPS Visit* 2:00 - 5:00	A. MIGLIORI *Resonant Ultrasound Spectroscopy* 1:30 - 4:30	FREE AFTERNOON	J. M. SABATIER *Porous Media* 1:30 - 4:30	T. J. MATULA *Sonoluminescence* 1:30 - 4:30	
DINNER 6:00 - 7:00	DINNER 6:00 - 7:00	DINNER 6:00 - 7:00	**BBQ AT PIT** 6:00 - 7:00	DINNER 6:00 - 7:00	**BANQUET** 6:00 - 7:00	DINNER 6:00 - 7:00	
INTRODUCTIONS AND ORIENTATION SOCIAL 7:30 - 10:30	DISCUSSIONS AND SOCIAL 7:30 - 10:30	DISCUSSIONS AND SOCIAL 7:30 - 10:30	OPEN EVENING FOR INDIVIDUAL ACTIVITIES OR INFORMAL DISCUSSIONS	DISCUSSIONS AND SOCIAL 7:30 - 10:30	DISCUSSIONS AND SOCIAL 7:30 - 10:30	WRAP-UP & GRADUATION 7:30 - 8:30 BONFIRE SOCIAL AT THE PIT 9:00 - 11:00	

SPONSORED BY THE OFFICE OF NAVAL RESEARCH AND ORGANIZED IN COOPERATION WITH THE ACOUSTICAL SOCIETY OF AMERICA AND THE NATIONAL CENTER FOR PHYSICAL ACOUSTICS AT THE UNIVERSITY OF MISSISSIPPI

CONNECTIONS IN PHYSICAL ACOUSTICS
AN INTRODUCTION TO PASS 2000

Anthony A. Atchley
Pennsylvania State University

DR. ATCHLEY: The abstract for this lecture is largely a disclaimer that I cannot cover physical acoustics or even a significant part of physical acoustics in three hours. Having confessed this, I have chosen the following approach. We will begin with a simple piece of acoustics and follow a path, adding more complexity and more physics as we go, and see where that gets us. In the process, I want to try to touch on at least most of the other topics to be addressed by the other lecturers.

[Transparency 2]

This diagram depicts a generic experimental system. Let's look at it and see where it leads us. Although one might quibble with the names I've assigned to its various parts, the elements of this system are common to almost every experiment. There is a source of a signal that runs through an amplifier that is tied to some acoustic source. There is also some type of receiver and a receiving system. There is also a test chamber that contains something to be studied. To simplify matters, as physicists do, let's put a perfect vacuum in there as a place to start. Later we can start throwing things into the chamber.

Suppose you set this experiment up. Set this frequency to a kilohertz. The source moves back and forth. What do you think I am going to detect? You would think nothing, right? There is, after all, a vacuum in the chamber and we all know that sound can't propagate through a vacuum. Are you really going to hear nothing coming from the loudspeaker?

First, there are plenty of reasons why you will hear a kilohertz coming out of here. There could be electrical pickup, cross talk. There could be structure-borne coupling. There could be some acceleration sensitivity in the receiver. In fact, you'd have to work hard to get rid of all these problems.

But suppose you get rid of all those problems? What are you going to hear? Buzzzzzz -- noise, right? The question is, is that nothing or is that something? Is there any information you can gain about the system from that noise? That is what Dr. Gabrielson is going to talk about. (By the way, I go first and then they have rebuttal time.)

[Transparency 3]

So let's build up from next to nothing. Let's put a single atom in the chamber. The question is, if you had the same system with a single atom or molecule in chamber and the source was oscillating at a kilohertz, what would you hear? What would you hear from that system? Would it be like "pong?" That is, the atom rattles around in the test chamber and finally hits the receiver that produces a blip? The question is what would the acoustics of a single atom or gas molecule be like? It turns out this problem is too hard for me, but Mo Greenspan wrote a paper, I guess, what? 30-some years ago in physical acoustics that told us about propagation of sound in gases at very low pressures. This is something I think may be worth discussing some evening.

[Transparency 4]

Let's add a lot of atoms or molecules to the chamber so we have a continuum. Now we've built up to something. Now what I want to know is how does describe the wave phenomenon in that chamber? Where do you start? How do you add more physics into it? Once you do that, what can you learn from it? What can you learn by adding pieces to the system? Answering these questions is the goal of this lecture.

[Transparency 5]

The place you begin is with the fundamental equations for a lossless fluid continuum. We have enough material in our chamber that it can be treated as a continuum, a fluid in particular. I am not going to talk about solids very much. We are also going to assume it is lossless, to start with -- we are going to put losses in later.

So what do you make of this? How would you interpret these equations and what can you do with them? You begin with the general equation for conservation of momentum. The Greek letter ρ, with no subscript, is the total density of the fluid, the ambient density plus the fluctuating part. u is the vector velocity of the gas. P is the total pressure, again the ambient part plus the fluctuating part.

The question I ask you is what do these equations really mean? What do you gather from the equation for conservation of momentum. What does it tell you? What concepts does it relate? It is really Newton's second law per volume. It relates net forces on pieces of the fluid to the acceleration of the fluid.

What about the equation of continuity? What does that tell you? It is coservation of mass. It says if I have a little segment of this gas I look at how much mass goes into it and how much comes out of it and the difference has to be what stays in there. These are the starting points for all of our work.

Normally, you would now linearize these equations and find the 1-D form. These are the standard acoustical approximations. However, introducing them does throw in some limitations for us. However, I will mainly talk about linear acoustics, and I am going to talk about 1D propagation for the first 80% of the lecture. I'll mention a bit about nonlinear acoustics at the very end of the lecture.

We also need to throw in an equation of state, and I have used a standard linearized equation of state. B is adiabatic bulk modulus. By the way, the 0's indicate ambient conditions and the sub-1 on ρ means the first-order fluctuating component. The p is acoustic pressure amplitude and we are assuming no net flow. Normally what you would do at this point is combine these three equations and derive a wave equation. However, we are not going to do that.

[Transparency 6]

What I want to do, instead, is to pretend that I can do something in terms of computational acoustics, which really is a joke. Let's take the same linearized form of the 1D equations and convert them into their simple finite difference forms. All I have done is represent the gradients with a difference in pressure, say, at a point $n+1$ and a point n, and I have also assumed $\exp(j\omega t)$ time dependence. Yes, I use j. Get over it. U is volume velocity and S is cross-sectional area.

So it is a simple transformation to go from the linearized 1-D forms to the finite difference forms. What does this tell you? Notice that the equatinos are obviously coupled. It says that the pressure at one point depends on the pressure at a previous point and something that has to do with the volume velocity. Similarly, the volume velocity changes by some amount that depends on the pressure. Coupled equations.

[Transparency 7]

Now what I want to do is show a pictorial representation of what these equations mean. You may soon think that the bulk of this lecture is devoted to equivalent circuit representation. It really is not. I am bringing in equivalent circuits because you have familiarity with them and it is a pictorial view to keep track of the physics. You can think of the circuit diagrams as

Feynman diagrams for acoustics. They get more complicated as the interactions change -- if you do not like the term "equivalent circuits," think of Feynman.

Let's take the momentum equation. I want to define this thing called in inertance, *M*, which is $\rho_0 \Delta x/S$. The standard equivalent circuit interpretation of the finite difference equations is that you have a segment of a duct of length Δx and area *S*, filled with a fluid. If you try to squeeze volume velocity through the segment, it will is going to result in a pressure difference. The electrical analogy of this process would be an inductance. Depending on the inertance and the volume velocity you build up a certain pressure difference across this element.

What does the pressure difference depend upon? Why would it depend on the cross-sectional area, for instance? When you look at these equations, I do not want you to just stare at them and say, ugh!, an equation. I want you to look at them, think about them, and see if you understand what they are telling you. So somebody explain why, as the cross-sectional area goes up or down, this pressure difference would change -- intuitively?

PARTICIPANT: [Inaudible]

DR. ATCHLEY: Yes, good! If the area decreases, you are trying to squeeze the same mass through a smaller area. You would think that should build up a bigger pressure difference.

That does it for one of our fundamental. I want you to keep this model in mind -- inertance.

As we go through this lecture we are going to add complexity, losses for instance, to this system. As we add complexity, we will build on our model. This is our starting point. I am hoping that at least the first part of this lecture -- actually, I am hoping for all the lecture -- there will be nothing startlingly new. What I hope might be different is the context it is put in and the connections you make with it.

[Transparency 8]

Now let's do the same thing with an equation of continuity. It says that if I put some volume velocity in one end and get volume velocity out of the other, those two do not necessarily have to be equal. Some mass can stay inside the segment. The difference in volume velocity is expressed in terms of this thing we call compliance, $C = S\Delta x/B$. *B* is the adiabatic bulk modulus.

Pictorially, think of this as electric current. If you put current in and you do not get the same amount out, that missing current must have gone somewhere. Electrically it would be shunted to ground through the capacitance, which is the electrical analog of as compliance. In acoustics what has happened is that you are just building up pressure in the segment.

Why, when the bulk modulus goes to infinity, would the compliance go to zero? What does infinite bulk modulus imply?

PARTICIPANT: Incompressibility.

DR. ATCHLEY: Incompressibility, right. If the fluid is compressible, then whatever you put in has to come out. Good.

[Transparency 9]

The next step, of course, is to combine these two concepts into one, because, in reality, any piece of this fluid is going to have both the property of inertance and the property of compliance.

So you have pressure at one end and volume velocity going in. You have a different pressure and volume velocity on the other end. What happens in-between is a combination of inertance and compliance. These effects are represented by combining the circuit elements as shown.

Suppose I have an initial pressure and an initial velocity. The compliance of the gas, causes some current, if you will, to go into ground, which modifies the volume velocity. Now you have a different volume velocity going across the inertance. The volume velocity going through the inertance results in a pressure drop.

Does it matter in which order I take these things into account? Let's arbitrarily put the compliance first and the inertance. Does it matter? It seems as if it could. For instance, if I have no initial pressure, then there will be no change in volume velocity due to the compliance. However, if we were to put the inertance first, then the volume velocity would generate a pressure difference and the compliance would have an effect. So you might imagine that it matters a little bit which order I take these things in. This is a question we want to try to get to, how to choose to order these things best, for instance, if you want to do numerical calculations.

[Transparency 10]

Now that we have the combined circuit representation, then let's see what we can calculate and how well we can calculate it. In this example, I have taken a constant cross-section pipe of some length, and I have broken it up into a number of segments. I will represent each segment in terms of a number of the lumped parameters, inertance and compliance.

To be specific, what I have done is slightly different from what I have shown you on the previous page. I took half of the compliance and put it on the other side of the inertance in the circuit representation. It does not really matter that much. This is just how I chose to do it.

I have plotted the pressure as a function of position. I start at the left end with a pressure amplitude of 1 and no volume velocity. Also, I have chosen a frequency so that 1 m corresponds to half a wavelength. The different lines correspond to the different numbers of segments used to represent the meter-long pipe.

Suppose you represent the entire pipe with one lump? You would think that would be pretty lousy. It turns out that the result is not as bad as one might expect.

Now if one uses four segments, each 25 cm long. You would think that might be way too big to do any good. It turns out the x's are the pressure predictions based on four segments. By the way, remember that I know things only at ends of the segments. The lines are just connecting the dots. The solid line is the exact result, a cosine curve. Four is really not bad and, depending on what you wanted to do, that might be good enough. Finally I pushed used 50 elements, so each is a 2-cm-long section of pipe. It does a pretty good job.

In effect what we have is a first attempt at computational acoustics. All you have to do is take the fundamental equations, discretize them, and write some eight or nine lines of code to calculate how variables changes across the elements. Pretty simple and we can get reasonable answers!

What else can you get now? I showed you a plot of pressure. You can just as easily plot the volume velocity. If you have pressure and volume velocity, what else can you get? Everything there is to get. You can get intensity. You can get power flow. You can get impedance. All good things flow. Just by taking the 1-D equations and doing the simplest possible finite difference interpretation.

By the way, everything I am presenting in this lecture is something that I have been able to calculate. This is an important point. Sometimes people will show you things and say, "Oh, you can do this," but what they really show you is something that is made up, a cartoon. I have not done that here. All the graphs you are going to see are the results of calculations I have done. The main purpose is to show you that if I can do it, you can do it. With Mathematica or MATLAB these days you can do a lot of things that I could not do as a graduate student.

[Transparency 11]

How far can you push it?

The upper left graph is the same one from the previous page. The other two graphs show how things degrade as the pipe gets longer. The upper right graph corresponds to a 2-m long pipe. The lower graph shows a 3-m long pipe. I thought it was interesting that even when the pipe was 2-m long, so the segments for the 4-element case are half-a-meter long, 4 segments does not do too badly. Obviously, 50 elements is better, and as you keep making things longer and longer, you can start to see more and more discrepancies even with 50 elements. You quickly get to the point that if you want to do what Ken Gilbert does and propagate kilometers in the atmosphere, this is not the best way to do it.

The question now is what would be the next-best guess for the next higher order approach to these kinds of computational problems?

[Transparency 12]

In the process of preparing this lecture and talking to colleagues, I noticed the following. Suppose you take the simple 1-D form of equation of continuity and conservation of momentum, and just rewrite them in the forms shown in the first two lines. Compare these forms to those used in, for example, fourth-order Runge-Kutta techniques. The forms are exactly the same. One might interpret, then, that these higher order numerical techniques are just a better guess of how to organize inertances and compliances to be more computationally efficient. We talked about how it might matter in what order you arrange the circuit elements things, depending on boundary conditions. Maybe the interpretation of higher-order numerical techniques is just a more general approach to arranging the circuit. Of course, with these higher order techniques you can make bigger step sizes and achieve higher accuracy. The point is, even if you go to something higher order, it still has the flavor of a simple model and the extension is not too difficult.

The main point of the first part of the lecture is that one can start with the fundamental equations, discretize them in very simple ways, and get numbers out. If the simple lumped parameter discretization does not yield satisfactory results, then going to a higher order numerical approach is conceptually no different. It is just a slightly different interpretation of the circuit elements that we used to represent our fluid.

By the way, notice that we have not done anything about the wave equation, I have not even mentioned it yet (to any extent). To do acoustics, there is no requirement to use a wave equation, necessarily.

We can go only so far sometimes by breaking a system up into discrete elements. The next step complexity is to apply the fundamental theorem of differential equations, that is, assume you know the answer and use it.

[Transparency 13]

Let's assume we are looking for wave-type. Returning to the differential form of the fundamental equations, we can divide the inertance and compliance by the length of the segment. M_ℓ and C_ℓ are inertance and compliance per unit length.

Seeking solutions of the form, exp($-jkx$), and substituting them in the fundamental equations, allows one to write the equations in matrix form. The values of k that satisfy these equations is then found by setting the determinant equal to zero. Doing so yields that for this simple system, $k = \pm\omega/c$, where c is defined to be one over the square root of the inertance per length times compliance per length. This relationship is the dispersion relation. Substituting $M_\ell = \rho_0/S$ and $C_\ell = S/B$ into this result shows that c turns out to be equal to the square root of bulk modulus over density. This is the adiabatic sound speed,. exactly as one would expected.

It turns out that sometimes getting this dispersion relation is all we really need. That is, we do not necessarily care about the behavior in detail of pressure and velocity. We just want to know the characteristics of the propagation.

By the way, why is there a plus or minus sign? Why are there two values of k? The interpretation certainly is a wave going one way or a wave going the other, but why are there two?

DR. WAXLER: Two linearly independent solutions.

DR. ATCHLEY: Thank you. Had we gone far enough to derive the wave equation, then you would all have resoundingly shouted "because it's a second-order differential equation," so you would have expected two roots.

Knowing the dispersion relation is going to be important for Dr. Keolian's and Dr. Sabatier's lectures. Therefore, I want to take some time to set up their talks, or perhaps it is better to say that I want to set them up.

So far we have explored simple lossless systems – the simple stuff. We began with the fundamental equations, and discussed their interpretation in terms of lumped parameter circuit elements. The use of these circuit analogies provides a way of holding the physics in our minds. We saw how these elements can be used to calculate pressure and velocity, or anything else one cares to know about an acoustical system. And we have just seen that if one does not want to take a computational approach, one can instead assume solutions and get a dispersion relation. The dispersion relation can tell you a good bit about the physics underlying a problem.

We are going to now go back to the beginning and follow exactly the same steps, but now we want to put in some realities of the gas. We're going to introduce viscosity and thermal conductivity and see how that modifies the fundamental equations and our interpretation of the them. After a while we will find the propagation constant, including viscous and thermal wall losses in our little segment. First, we introduce the physics.

[Transparency 14]

I give credit to the approach taken in this part of the lecture to Greg Swift and everybody else in this room or not in this room who has taught me anything about thermoacoustics. The approach I am taking is a standard approach used by thermoacousticians. Dr. Sabatier will discuss much the same phenomena with slightly different parameters, but it is the same physics.

What I want to do now is introduce viscous effects into our fluid. Acoustic attenuation in the bulk of the fluid is not going to be important for the systems we deal with in this lecture. The dominant dissipation in these systems is due wall losses, the only kind of attenuation I am going to include.

If one wishes to take into account viscous effects in the viscous fluid, how will that modify the fundamental equations with which we started? What the new equations must we deal with? The answer is the Navier-Stokes equation, that is, the simple force equation. Newton's second law per volume shows up again, but now we have to tack on terms to include the effects of viscosity. You can go to Landau or Lifshitz, or somewhere else, and pull out the equation shown at the top of Transparency 14. The left-hand side and the first term on the right-hand side make up the equation for conservation of momentum we what we used at the beginning of the lecture. All we have done is tacked on the two terms to handle viscosity.

By making "reasonable" assumptions, the N-S equation can be simplified quite a bit. What are those reasonable assumptions? I am going to be a bit glib -- read Swift's review article in thermoacoustics and you find out what "reasonable assumptions" mean. For us it means we are going to linearize things. We are going to throw away terms that are just too nasty to deal with. And, of all the various gradients and divergences that matter to us, the gradients in the transverse direction are going to be the most important. This is because the transverse gradients are going to be dependent on a quantity called the penetration depth, which typically, for our purposes, is a 10th of a millimeter or so. For the longitudinal gradients, the characteristic scale is on the order of an acoustic wavelength, and for all the cases we deal with that length is much longer than a penetration depth. This means all the derivatives associated with them are much smaller.

If we assume propagation in a circular pipe or circular duct, the N-S equation reduces to the second equation on T14. Let's identify all the terms. The left-hand side is equivalent to the time derivative of u. The first term on the right-hand side is the linearized pressure gradient. Te final term comes from keeping only the radial part of the gradient. Go back and look in your notes and convince yourselves that all we have done is add this radial gradient term to the equation for conservation of momentum. Notice that in the absence of viscosity that goes away and we're left with the same equation we had before. Introducing viscosity adds the last term. What do you think that adding this term is going to do? It is clearly going to add something else into our model. What I want to do now is figure out what that term does.

Would all the discussion leaders please raise their hands? I want the students to look around and see all the discussion leaders. Every single one of them will be glad to tell you how you go from the second equation to the third. (Laughter)

The third equation is the solution to the differential equation given just above it. J_o is the spherical Bessel function of order zero. δ_v is the viscous penetration depth. If you are not familiar with it, I will say a few words about it in a few minutes. μ is viscosity; ω is angular frequency; and a is the radius of the duct.

One thing that has happened now that we've introduced viscosity is that things are no longer uniform across the radius of the duct. There is now radial dependence. We'd like to get rid of that as soon as we can. Therefore, the next step is to integrate the velocity over the cross-section of the duct to get volume velocity.

[Transparency 15]

If one integrates u with respect to r -- it is not too tough – one gets the result at the top of T15 for volume velocity. Comparison the result we had earlier shows that the relationship between volume velocity and pressure gradient has been modified by the factor f_v. This function is going to show up in a lot of places. Dr. Keolian will show you similar functions for thermoacoustics. Dr. Sabatier will show you similar functions for porous media, except you are not going to use different notation, right?

DR. SABATIER: This time I am going to use the viscous penetration depth. (Laughter)

DR. ATCHLEY: The point is that by adding viscosity into our model now we are starting to make connections with thermoacoustics and porous media work. We will see where that gets us.

[Transparency 16]

This graph shows f_v as a function of a/δ. There are three curves. The solid line is the real part of f_v, the dotted line is imaginary part, and the dashed-dot line is the magnitude. The horizontal axis can also be interpreted as a frequency scale, because the penetration depth has an sqrt($1/\omega$) dependence.

In cases where the radius is large compared to the penetration depth, maybe a centimeter or more in our case, f_v approaches a high-frequency limit. The dissipation function get smaller. When a/δ_v is small, that is, when the penetration depth fills the cross-section of the pore, the imaginary part is zero and the real part is magnitude of one. There is a peak in the magnitude of the imaginary part of f_v when the radius is on the order of the penetration depth.

Dr. Wilen and his students measure these functions for a living and I am sure they would be glad to tell you about the intricacies of doing that. This function and others related to it show up in a lot of places. It will be important to remember that something happens when the size of the pore is comparable to the penetration depth. You may want to keep that in mind as we go through the lecture.

[Transparency 15]

Back to our finite difference version of the equation. If f_v were zero, the third equation on T15 would be the same as the lossless version we have seen before. Introducing viscosity modifies the last term. If you compare it with the lossless version you will see that we modified the inertance by that factor, $1/1\text{-}f_v$. That modified inertance, because f_v is complex, has both a real and an imaginary part. What do you think that is going to do?

Before, the inertance showed up in the representation as a reactance. There was no loss associated with it. However, now the inertance is complex. Therefore, we will now have a part that will lead to a circuit element having a real impedance, that will lead to dissipation. It is easiest to look in what is called a boundary layer limit, in which a/δ_v is a large number.

[Transparency 17]

If one go through some algebra, one find that in the boundary layer limit, the ratio of J_1 to J_0 is j. Let M' be the modified inertance including viscosity. In the boundary layer limit you can see how it breaks up into a real part and an imaginary part. The real part is just (lossless) inertance we had before. The lossless inertance also shows up in the imaginary part but modified by the factor δ_v/a.

[Transparency 18]

After a little math, one finds the boundary-layer-limit equivalent resistance R_v due to viscous interactions between the fluid and the wall of the duct. In the lossless case, the inertance led to a pressure drop across the segment of duct that was 90 degrees out of phase with the pressure at the entrance. Now we have a term that is in phase with it. The electronic analog to that is resistance; that is, the pressure drop is equal to the current times the resistance.

[Transparency 19]

Introducing viscosity leads to a modification in our equivalent circuit representation of our duct. Volume velocity shows up in both terms in the finite difference equation, so, electrically, that is analogous to circuit elements in series since they share the same current. So the inertance is modified slightly because there is viscosity and there is now this new term for resistance.

[Transparency 20]

A graph of M'/M is shown on T20. The important thing to notice is that whereas the real part has very little dependence on a/δ_v, the real part gets very large for values of a/δ_v on the order of one and smaller. In other words, the resistance is very high in small diameter ducts.

[Transparency 21]

For those of you who are not familiar with the viscous penetration depth, let me give you a mini-tutorial on it and then we will take a break. Suppose you had to solve the following problem. A viscous fluid oscillates parallel to a motionless rigid boundary. Infinitely far away from the boundary, the velocity amplitude is U_o. On the boundary, the velocity amplitude is zero. Find the velocity as a function of distance from the boundary and time, i.e., $u(y,t)$. That is, what is this profile of the fluid velocity in a viscous fluid? Dr. Keolian is going to show us an animation that has the real details of this envelope.

[Transparency 22]

How do you find the profile? You go back to the diffusion equation, which I have written at the bottom of T21, and you ask all the discussion leaders in the room how to solve it. Just kidding. When I solve it, I assume I know the answer. Assume an answer of the type given at the top of T22. Now you might be tempted to say, "You can't do that in a diffusion equation, that's only good for waves." Well, try it and see what you get.

If you substitute this result into the differential equation, you will be left with an algebraic equation that will tell you what values of κ make the assumed solution work. If you do that, you find that this κ has equal real and imaginary parts. The quantity sqrt($2\mu/\rho\omega$) pops right out and that is the definition of the viscous penetration depth.

If you break up this solution, you can see that it has a propagating part, with a wave number equal to $1/\delta_\nu$. There is also a decaying part that decays with the same scale length. I call it a mini-tutorial, one weeks of fluid dynamics in 38 seconds.

We will take a break and when we come back I want to introduce thermal conductivity in our model.

All right, introducing viscosity changed the momentum equation. Introducing thermal conductivity is going to change the equation of continuity. So how are you going to introduce thermal conduction losses between the fluid and the walls? You write down the general heat transfer equation make reasonable assumptions to simplify it.

When one linearizes this equation, terms including viscosity are discarded. Therefore, in a linear model one can separate viscous and thermal effects. However, at the nonlinear level these terms are retained. Does that mean at the nonlinear level you cannot separate those two effects? I leave this question for further discussion.

Now we ask the question what happens to the gas when thermal conductivity is included? When the gas inside the segment is compressed and heated up, then the pressure changes, the temperature changes, and the density of the gas changes.

The second equation on T23 is the linearized version of the first equation. It relates pressure changes, density changes, and temperature changes. If you go back and look at the equation that we were just talking about for viscosity, you will see that this equation has exactly the same form. Therefore, the solution is exactly the same; that is, we will have an f function for thermal conductivity. The third equation describes how the temperature varies with position in the segment. β is the thermal expansion coefficient of the gas. T_o is the ambient temperature and T_1 is the first order fluctuation in temperature. c_p is the isobaric specific heat capacity. K is the thermal conductivity of the gas. The term in the brackets has exactly the same form we found before for viscosity.

This equation describes how the temperature changes across the duct as a result of heat transfer to the walls. We still have to get to the equation of continuity. How do you do that? What concept is an equation of continuity? What parameters does it connect?

[Transparency 24]

Thermodynamics tells us how to relate changes in variables such as pressure, temperature, and density, as shown in the first equation. To find the linear acoustics version of this equation, let $d\rho$ becomes the acoustic fluctuating part ρ_1, dp becomes the acoustic pressure p, and dT becomes the acoustic temperature fluctuation T_1. The partial derivatives can be expressed in terms of properties of the fluid, namely, the isothermal bulk modulus B_T and the thermal expansion coefficient β. Using these relations, we can relate acoustic density changes to pressure changes and temperature changes, as shown. On the previous page, we had an equation for T_1, as a function of acoustic pressure. Now we can express it in terms of the acoustic density fluctuation.

[Transparency 25]

Just to show you that there is light at the end of the tunnel, the top of this page says we are almost there. The idea behind all of this is not to bore you with equations, it is just to bore you in general. (Laughter)

I am including all the details for completeness. I realize that you will not follow the steps here, but the details are there so that you can find them if you ever need to later.

After significant work you get down to the equation at the bottom of T25. This equation relates the gradient in volume velocity to the acoustic pressure. If we were to neglect thermal conductivity, this function, f_κ, which is exactly the same form as f_ν, would go away, leaving the lossless version of the equation of continuity

[Transparency 26]

Let's quickly get to the bottom line. That function f_κ includes the thermal penetration depth. If you go back to T21 and ask yourself what the thermal penetration depth is, it comes out of the following problem. Instead of a movable boundary and viscous fluid oscillating back and forth, suppose that up at infinity we have an oscillating pressure. Up at infinity you know what the temperature change is, right? It is the adiabatic temperature change, because infinitely far away there is no time for heat to conduct from up there down to the boundary. At $y = 0$, there is an infinite heat sink. Therefore, the temperature at that boundary will not change. Hence, the temperature fluctuation is zero there. You have exactly the same boundary conditions. What kind of equation describes thermal conduction problems? A diffusion equation, right? You go through the same process and you get thermal penetration depth, which has the same interpretation as the viscous penetration depth.

Remember when we included viscosity, the inertance of the fluid was modified. There is a graph on T20 that shows how the inertance changed with frequency or the size of penetration depth compared to radius. That graph shows that is a modification of the real part and the imaginary part. As the penetration depth gets large compared to the radius, the imaginary part zooms through the roof. It completely swamps the real part, which means the dissipation from viscosity dominates in small pores. Do you think the same thing is going to happen with thermal conduction? We are about to find out.

[Transparency 27]

The consequence of adding thermal conductivity to our model changes the compliance. C' represents the compliance modified by including thermal conduction. It differs from the losselss compliance C by the factor $1 + (\gamma - 1)f_\kappa$. Notice that the compliance is now complex.

[Transparency 28]

This page shows a graph of this complex compliance on T28. As before, the solid line represents the real part and the dashed line the imaginary part. The real part behaves similarly to the real part of the inertance. That is, it changes slightly as a function of a/δ_κ. Can someone explain the transition in the real part of compliance? By the way, this graph is for air. The real part of C' and it goes from 1 at large values of a/δ_κ to 1.4 for small values. Can somebody explain that change? As the thermal penetration depth gets small compared to the pore, what is going on? At large values of a/δ_κ the process is adiabatic. That is, at penetration depths that are really tiny compared to the pore, most of the gas cannot conduct heat to the wall in the time allowed, so it behaves adiabatically. Thus large values of a/δ_κ is the adiabatic limit.

For small values of a/δ_κ, the penetration depth swamps the radius and all the gas is in good thermal contact with the wall. Therefore, acoustic processes in this limit are isothermal. The transition the real part of C' is indicative of the transition from adiabatic propagation to isothermal propagation. γ is 1.4 for air.

Now what about the imaginary part of C'? Notice that its dependence on a/δ_κ is completely different from that of the real part. Further, it is completely different from the behavior seen in the viscosity-modified inertance, that went through the roof for small values of a/δ_ν. Here, notice that the imaginary part is very small at either limit of a/δ_κ. The acoustic processes are lossless in either extreme. Why? Adiabatic processes and isothermal processes are both lossless. It is only in the region where the pore is roughly the size of the penetration depth that significant losses due to thermal conductivity come into play. By the way, that little bump in the imaginary part of C' is thermoacoustics, at least one flavor of thermoacoustics. Dr. Keolian will tell us more about that.

[Transparency 29]

Combining everything leads to the finite difference version of the equation of continuity shown at the top of T29. As you can see, introducing thermal conduction leads to another resistance into our model, representing the lossy part. Notice that it is drawn as a resistance to ground. What happening is that the thermal conduction is modifying the compliance. It is taking volume velocity out of the flow. If you squeeze gas into the segment, the ordinary compliance leads to a pressure build up in the segment that reduces the volume velocity coming out. However, the thermal conduction from the gas to the walls leads to another reduction in volume velocity coming out.

It is important to remember that thermal conduction introduces this resistance. Can you imagine any circumstance under which this resistance could be negative, where it could act like a source of volume velocity? We'll leave that question for another lecture.

[Transparency 30]

The discussion of compliance on T27 and T29 were in the boundary layer limit, when $a/\delta_\kappa \gg 1$. T30 shows what the real and imaginary parts of the compliance look like when the boundary layer approximation does not apply. Other than that, there is nothing new on this page.

[Transparency 31]

Of course, a fluid element displayes properties of both inertance and compliance. This is borne out by examination of the finite difference versions of the equation of continuity and conservation of momentum shown at the top of page T31. These equations lead to a combined equivalent circuit representation as shown at the bottom of the page.

To summarize this part of the lecture, we have seen that by adding complexity to the fundamental equations, we can build upon our representation of the acoustic behavior of a fluid. We have expressed the physics involved with two parameters, inertance and compliance. We have seen how the finite difference forms of the fundamental equations are modified. If we wanted to at this point, we could return to the model used to generate the plots shown on T10 and T11 and introduce losses. However, rather than spend to time to do this, I want to return to a different topic that we discussed earlier, the dispersion relations.

[Transparency 32]

Here are the differential equation forms of the fundamental equations. Notice that they have been modified by putting primes on the inertance and compliance per length. The primes indicate that the viscosity and thermal conductivity-modified forms of these quantities should be used. As before, by assuming propagation-type solutions, these equations can be cast in matrix form. Also as before, the propagation constant is found by taking the determinant of that matrix. And as before, there are still two values, corresponding to waves propagating in two directions. We can write the propagation constant k as $\pm\omega/c_{\kappa\nu}$, where $c_{\kappa\nu}$ is the phase speed. Notice that the phase speed is now complex, because f_ν and f_κ are complex.

Let's look at the complex phase speed for sound propagating through a segment of duct. To do so, we are going to break it up into two parts. We are going to look at the thermal conduction part first and then we are going to look at the viscous part.

[Transparency 33]

The solid line is a graph of, using the left axis, the ratio of that complex phase speed including only thermal conduction losses divided by the ordinary thermodynamic (lossless) sound speed c_o as a function of, again, the same parameter, a/δ_κ plotted on a log scale. You will notice that in the high-frequency limit, the phase speed divided by the ordinary adiabatic sound speed is 1, which means we have adiabatic propagation of sound. As you transition through this boundary layer range you get down to a value of about 0.84 relative to the adiabatic sound speed for air. Why is it 16% lower? Is Sir Isaac Newton in the room? That is the isothermal sound speed. It is about 16% lower than adiabatic sound speed. The transition is, again, from adiabatic propagation to isothermal propagation. We saw the same behavior when we looked at the complex compliance but this representation, I think, is a little bit clearer. Notice that the action happens where the radius is comparable in size to the penetration depth.

The dashed line is the factor exp($-\alpha\lambda$). I have taken the imaginary part of the complex propagation constant to get the attenuation coefficient α. λ is the acoustic wavelength, which is $2\pi/\mathrm{Re}[k]$. The factor exp($-\alpha\lambda$) is the amount by which the amplitude of the sound will decay as it propagates a distance of one wavelength. In the high-frequency, adiabatic limit this factor is one. There is no loss for adiabatic propagation. At the low-frequency, isothermal limit, the factor is also one, again lossless. The propagation is lossy only in the intermediate range where the radius is comparable to the penetration depth. You see that in the worst case, as sound propagates one wavelength, it decays to about 60% to 70% of its initial value, so it decays about 30% over a wavelength.

[Transparency 34]

This graph shows the same thing, but this time only including viscosity. Comparing this graph to the previous one, they are very different. Dr. Sabatier will going to tell us why. The phase speed starts off in the high frequency limit equal to the ordinary sound speed. But, as the penetration depth gets larger and larger compared to the radius, the sound speed starts to drop, just as it did in the thermal case. However, in this case it does not level out as it did before, it plummets and goes to zero.

If you were to call waves moving in the high frequency limit fast waves, what would you call waves in the low frequency limit? Slow waves. Dr. Sabatier will tell us all about slow waves. The attenuation factors get really big, too. So as you can see thermal and viscous effects cause drastically different behavior in the propagation characteristics of sound in a duct.

[Transparency 35]

If you put them together, what do you get? That is the case shown in the next graph. AIt is not very different from the one on the previous page. As before, for small pores the phase speeds can get really slow and attenuation factors get really high. This raises an interesting question about ceiling tiles and sound absorbers. If you look in textbooks that talk about attenuation of sound in sound-absorbing ceiling tiles, they almost always neglect the thermal component of attenuation, they include only viscosity. When would this be acceptable? The choice of answers is, A, never; B, sometimes or C,always. The answer is B, when the pore is small in diameter compared to penetration depths. Go back and look at the thermal plot. There is no loss down here because it is isothermal. So if the pores are really small compared to the penetration depth, you can include just the viscous terms. The viscous terms dominate and that is why you can get away with it.

The point of this part of the lecture, is to show you that you can go through a lot of complexity and keep building up equivalent circuits and finite difference equations. However, but if you just want to know something about the propagation characteristics, get the dispersion relation. It tell you quite a bit about the nature of the propagation.

Everything I have done so far has been with a circular pore. A lot of things in thermoacoustics is done with other types of cross-sections. The only thing that needs to be modified substantially to take these different shapes into account is that there will be different equations for the f_κ and f_ν. However, the frequency dependences of the different f's are all more or less the same.

The physics we've discussed so far is relevant not only to thermoacoustics, but also to porous media. There is one big difference, however. Rhe ground is not made of a bunch of straight, circular pores. The pores in the ground bend and the cross section is certainly not constant. The other thing we have to worry about is what is dirt? Excuse me, what is soil? Something you can make acoustic measurements in and make Dr. Sabatier happy. (*Laughter*) There is a solid part and there is a gaseous part. If you want to couple sound into the ground, how are you going to get it in there? There are a couple of ways, right? You could push on the solid part and get that moving or you could push on the gas part and get that moving. If you push on the gas part, you talk about the propagation in the gas. That is what we have discussed. When you push on the solid part, do you think that the wave you generate will be faster or slower than when you push on the gas? This is the origin of the so-called fast wave, propagation through the solid part of the ground. The slow wave is propagation through the gas. However, recall that his is a viscous medium. As the gas sloshes in and out of the ducts, there will be drag on the duct wall. Or if I move the solid part, the gas will get dragged along. So there is coupling between the two modes of propagation. You can couple between the sound in the gas and the sound in the solid. As I have said, Dr. Sabatier is going to say something about this.

[Transparency 36]

I asked a question a few minutes ago: Are there any circumstances under which the resistance due to thermal conduction could be negative? The answer is yes. If you start with fundamental equations as we have done but include the possibility that our duct had a temperature difference along the axis, you can derive, again, the equation of continuity. However, now the results will be different, as shown at the top of T36. The difference can be represented as an additional equivalent circuit element, as Greg Swift and colleagues have shown us. The new element, represented by the diamond, has the properties of a current source that depends on the the temperature gradient. If there is no gradient, we have the equation we had previously. If there is a positive gradient, the factor *e* will be positive and the element is a source of additional current. If the gradient is negative,*e* will have a negative value ad the element will be a sink.. Therefore, depending on the magnitude and sign of the temperature gradient, current will either be added or taken away.

The point is that even introducing the complexity of a temperature gradient can be represented by an equivalent circuit element. The equation at the top of T36 is not so easy to derive. That is why I say in quotation marks, "it can be shown." It can be shown but I am not going to show it.

We are now going to abandon this equivalent circuit approach and concentrate more on propagation constants.

This is a good time for a break.

DR. ATCHLEY: I knew the second part of this lecture was going to be brutal. However, I wanted you to see where the effects of real gas properties, such as thermal conduction and viscosity, come in and how they modify the fundamental equations. I did not know any other way to do it than just to be brutal about it and drag you along. I hope the notes are complete enough that if you really want to understand it you can go back and find all the details. It is really more of an historical record for completeness. I suspect many of you are halfway burned out. I apologize.

I think the next hour is going to be better, because now we have built up all physics that we need for understanding bigger systems. We could construct a larger system with equivalent circuits. However, we'll take a different approach. Rather than use a whole lot of circuit elements, we will take advantage of the fact that we are seeking wave-type behavior. We will construct our system from building blocks. Each block can be arbitrarily long compared to the wavelength. The only requirement is that the individual building blocks have uniform properties within them.

[Transparency 37]

If that is true, then we can treat segments in terms of counterpropagating waves because we know how to calculate the propagation constants. That is, we know the dispersion relation. So here is the approach we are going to take now.

Suppose you know the pressure and volume velocity at end of the system, or equivalently, you know the impedance at that end. Typically we want to know what the acoustic parameters are at the input to the system. To get this information, we'll use a "propagation matrix" approach. Here is how it works. We have two counter-propagating waves. We know how to calculate the coefficients in terms of boundary conditions.

We have M segments. We start at x_0. The ends of the subsequent segments are denoted as x_1 through x_N. Pressure and volume velocity are expressed as shown. Also, each segment can be characterized with a characteristic acoustic impedance Z_n.

[Transparency 38]

Assuming that pressure and volume velocity are continuous, the boundary conditions at the end of the last segment can be expressed in the form shown at the top of T38. Assuming that the M-matrix can be inverted, then this equation can be solved for the matrix C_N, which contains the propagation coefficients for segment N.

[Transparency 39]

Now we move to the next junction, between segment N-1 and N. Application of continuity of pressure and volume velocity leads to the equation at the top of T39. This equation can be solved for the matrix C_{N-1}.

[Transparency 40]

You finally arrive at the coefficient matrix C_1, which contains A and B in the first segment. If you know A and B in the first segment, you know the pressure and volume velocity at the beginning. Problem solved.

This technique allows you to break a complicated system up into a bunch of segments having different cross-sections, different lengths, even different gases -- it does not matter -- as long as the properties within any one segment are uniform. Let's apply the technique to a system to see what we get.

[Transparency 41]

Consider a pipe one inch in diameter, one meter long with a rigid termination at the end. The ratio of P to U at the rigid end is known. The question is what the ratio of P to U at $x = 0$? I have arbitrarily broken the pipe up into five segments, some having different lengths just to illustrate the flexibility of the technique. The graph at the bottom of T41 shows magnitude of the ratio of the pressure to the volume velocity at $x = 0$ as a function of frequency. Not surprisingly, you see that the impedance peaks at frequencies corresponding to a half wavelength in the pipe, a full wavelength, and so on. Keep in mind that the model includes thermal and viscous wall losses, so the quality factor can be determined from the width of the peaks. The circles indicate the peak frequencies

Why would you expect the peaks to decrease in amplitude with increasing frequency? Is it just an artifact of our calculation? Who knows if it is real? How do you check it out? That is a pretty striking property that these amplitudes decrease in some monotonic way. There must be a simple way to figure out what should happen.

Let me ask you this. This is a pop quiz for all the discussion leaders. If you are working from the rigid end towards $x = 0$, how does the pressure vary with x? It varies as cosine. What about volume velocity? Sine, right? So how will the impedance vary? If you dig back far enough, you will remember that the input of mechanical impedance is something like $-j\rho_o cS \cot(kL)$? What is kL for the first peak? π. What is the cotangent of π? Infinity. What is the kL for the second peak? Do I hear 2π? What is the cotangent of 2π? Infinity.

I suckered you into it. What is kL at the peaks? Remember k is complex. So kL is largely π or $N\pi$, but not exactly. There is a small imaginary part that provides attenuation. That decrease in amplitude of the peaks in the impedance at the input is due to the frequency dependence of the attenuation in the pipe. Why should it be frequency dependent? What is changing as the frequency changes? The penetration depth. How? It goes as the square root of frequency. If you break kL at the peaks into real and imaginary parts, you'll find that the real part that is close to a multiple of π. Throw that away. If you take the cotangent of the complex part, you will get exactly the ratios that come out of here. I have done it.

This discussion brings up an interesting sideline. The attenuation coefficient goes up with the square root of frequency in this system. You would normally think that higher attenuation would mean lower quality factor Q. However, it turns out the Q actually goes up as well. Both attenuation and the Q go up. Have you seen an electrical circuit where that is true, when you add more damping and the Q goes up? I think it is a parallel LRC circuit.

[Transparency 42]

DR. ATCHLEY: Now what I want to do is change one of the segments and see what happens. Suppose the fourth segment has a radius half the value of the rest of the pipe. This situation is shown on T42. Compare the graph with the previous one. It is quite a bit different, at least for certain modes. Some modes are not changed as much as others. Professor Garrett will show us an example of how this behavior can be applied to generate very large amplitude standing waves in resonators.

Now, suppose I set up this experiment in a black box. Also suppose that I told you that the pipe is one meter long, how many segments there are, how long they are, and the diameters of the four equal-diameter segments. Finally, I tell you what the peak frequencies are -- not the values of the impedance at the peaks, just the frequencies. What could you do with that? It turns out you can tell me the diameter of fourth segment. That is, you can tell me something about this system from a measurement of the input impedance. That is illustrated on the next graph.

[Transparency 43]

I did give my computer program the frequencies from the previous graph, not the impedance, just the frequencies. I told it that the fourth segment had a diameter somewhere in the range between the original value and one-tenth of that value. I asked it to find the radius of the fourth segment.

The computer came back and said, "I think the best fit is 1.269 cm." Not too bad. I then took that number and recalculated the response of the system. The result is shown in the graph. The x's make the peak frequencies. The circles are from the graph on T42. I think the match is pretty good. The point is that by measuring the resonance frequencies of the system and by knowing something about the system you can calculate, for instance, how large or small that constriction was. Keep in mind that our model is one-dimensional. Therefore, we have neglected complications due to abrupt changes in cross-section.

This part of the lecture was intended to set up Dr. Migliori. The main point is that by measuring properties of a system and a knowing something about the system, say, the eigenmodes or eigenfunctions, it is possible to calculate other properties of that system. In my case, I calculated the radius, which tells you something about the compliance and the inertance.

You could extend that to elastic constants if you went to a solid. Dr. Migliori is going to tell us how to do this on more complicated systems and what you can learn about the physics of those systems.

[Transparency 44]

Let's ask ourselves how you might include propagation in more than one direction in physical acoustics. A place to begin is with the Helmholtz equation. If you take the ordinary linear acoustic wave equation, for instance, in pressure, and you take the temporal Fourier transform, you get the Helmholtz equation. You can get the Helmholtz equation by substituting single frequency solutions into the wave equation. The Helmholtz equation is the starting point for this part of the lecture. Instead of discussing pressure as a function of position and time, we will be dealing with the temporal Fourier transform of pressure, which is going to be a function of position and frequency. Even though P is the Fourier transform of pressure, I am going to refer to it as pressure for simplicity.

[Transparency 45]

What follows is based on things that I've learned from the fields of Fourier optics and Fourier Acoustics. By the way, there is a great reference for this on the market now by Earl Williams from the Naval Research Lab called Fourier Acoustics. .

Suppose we take the spatial Fourier transform of the Helmholtz equation. -- Why? Because we can. Bear with me. -- Let's do it in rectangular coordinates so we have an explicit expression for the Laplacian. Take the Fourier transform in the x- and y-directions, that is, the transverse Fourier transform.

[Transparency 46]

The result is the following: After you take the Fourier transform, the derivatives in the x- and y-directions become multiplications by the transverse wave number components k_x and k_y. Therefore, the Helmholtz equation has the form shown in the first equation. Using the relationship among the various wave numbers given on T45 , the transformed Helmholtz equation can be written in the form shown in the third equation. You all recognize this familiar differential equation. It is the 1-D Helmholtz equation. And, you all know what the solutions are, counter propagating plane waves. What good is this?

[Transparency 47]

Consider the following situation. Suppose you only have waves propagating in one direction, say in the positive z-direction. This is an important assumption. The situation is much more complicated if you have waves going in both directions. Also, suppose you know the pressure in some plane, say at $z = 0$. Take the spatial Fourier transform in that plane. According to our one-way propagation assumption, the transform at $z = 0$ equals the coefficient A. How would you find the pressure here in some other plane at $z > 0$? Planar Fourier acoustics says you do the following. To find the pressure in some other plane, propagate the transformed pressure in the $z = 0$ plane forward simply by multiplying it by the phase factor $\exp(-jk_z z)$. This process gives you the spatial Fourier transform of the pressure in the new plane. To find the pressure in that new plane, take the inverse spatial Fourier transform. You're done. Let's do an example.

[Transparency 48]

Suppose the pressure amplitude in the $z = 0$ plane is given by the function shown in the top graph. It is zero everywhere except across some finite width, where it is one. Given this pressure distribution, find the pressure in some other plane at $z > 0$. The first step is to take the spatial Fourier transform of the pressure distribution. What is the Fourier transform of this function? A sinc function, as shown in the second graph. In principle the range of the sinc function is from plus to minus infinity. I just stopped because my page was of finite size.

Now, pick a point on this graph. It corresponds to a specific value, if this is a 1-D problem, of the value of k_x. What does a specific value of k_x mean in a Fourier representation? Remember, this rectangular pressure distribution is made up of an infinite number of sinusoids. The positive wave numbers k_x propagate in the positive x-direction and the negative values propagating in the negative x-direction. A particular value of k_x corresponds to a constant amplitude, infinitely-long sinusoid propagating in the x-direction with that wave number.

Describe the radiation from this wave. It is a plane wave propagating in a direction determined by k_x and k as shown in the figure in the lower left corner of T48. Keep in mind that for a constant frequency, k is constant. Therefore, each different value of k_x corresponds to a different direction. Also, each different value of k_x corresponds to a different value of k_z. Therefore, each different k_x-component of the transform advances in the z-direction with phase, given by $\exp(-jk_z z)$.

Notice when k_x = zero, the direction of propagation is $\theta = 0$. Suppose $k_x = k$? What is the direction of propagation? 90°. The range of k_x from 0 to k corresponds to propagation angles from 0 to 90°. This range is indicated by the rectangular window in the second graph. In the third graph, this range of wave numbers has been converted to angle. There is a one-to-one correspondence between each value of k_x in the window and an angle from –90° to +90°.

What happens to values of k_x greater in magnitude than k? To what angle do they correspond? No real angle. They correspond to evanescent components of the field. It is only the values of k_x within the range $\pm k$ that survive in the farfield of real. The wave number components outside that range correspond to nearfield effects.

An elementary explanation of a null in the farfield radiation from sources is destructive interference of the pressures propagating from different parts of the source. The Fourier acoustics interpretation of a null would be that the Fourier transform is zero for that particular value of k_x. Therefore, the pressure propagating along that particular angle has zero amplitude.

This is going to be important for Dr. Gilbert's lecture, particular wave numbers correspond to particular angles of propagation.

DR. HAMILTON: You get a nice benefit you get from this. In the farfield of any directional radiator the beam pattern is the Fourier transform of the source point.

DR. ATCHLEY: Thank you. I meant to say that: The farfield radiation pattern is the Fourier transform of the source function shown in the first graph. It also explains why the nearfield and the farfield are different, as shown on the next slide.

[Transparency 49]

This slide is Fourier acoustics on one page. Suppose you have a wave normally incident on a rigid baffle that has a slit in it. What does a diffracted field look like? The first step is to find the pressure in the plane of the baffle. The standard Kirchoff approximation is that the pressure is given by a rectangular function. This assumption, of course, neglects edge effects.

So the pressure in the baffle-plane is known. Take the spatial Fourier transform. It is a sinc function. Propagate the transformed pressure from 0 to z according to the plane wave propagator $\exp(-jk_z z)$. Keep in mind that k_z depends on k and k_x. Some of the Fourier component will have real values of k_z, others imaginary. As the wave propagates, the Fourier transform evolves. Fourier components with real values of k_z simply change phase, while those with imaginary values of k_z decay. As the wave propagates farther and farther, these components decay more and more. In the farfield they would be completely gone. Once the transform has been propagated the appropriate distance z, take the inverse spatial Fourier transform to get the pressure in that plane. This is the modern way to do diffraction problems. If you buy into this wave number approach, it is pretty powerful.

By the way, this is also the foundation of nearfield acoustic holography. You measure the pressure distribution in one plane, you take the Fourier transform, you propagate it to another plane and you can reconstruct the field. If Dr. Maynard were here, he would tell you that it is not necessary to worry about the evanescent behavior for any finite source, because instead of representing the field in terms of plain waves, you can represent it in terms of spherical harmonics.

So where are we? Almost finished, you will be glad to know. Let me summarize what we have done and then see where we have yet to go. We started by assuming simple 1-D propagation and we found techniques that will allow us to understand those systems. One was through a simplistic finite-difference approach that at least allowed us to open the door to computational acoustics, although that door can open a lot wider. The finite-difference equations led us to lumped parameter equivalent circuits. We found out that by understanding the parameters, you can determine the propagation constant, the complex wave number. From that you get the phase speed and attenuation coefficient.

Next we introduced more complexity, such as viscosity and thermal conduction. We found that, at least in the linear world, introducing viscosity modifies the momentum equation and introducing thermal conduction modifies the equation of continuity. Even though it added complexity to the system in more ways than one, this added complexity could still be expressed in terms of inertance and compliance. We also found that we could still get the complex propagation constant in a relatively simple way. This led to a discussion how phase speed and attenuation frequency.

We used the propagation constant to go beyond what you can do in small differential elements, and talked about propagation in a system that can be broken up into subsystems that can be long in terms of the wavelength. We used this approach to look at, for instance, the impedance in a simple 1D system. We saw that if you change the properties of one of the subsystems, for instance, the radius, that changes the resonance structure altogether. You can use the technique for that purpose alone or you can use the measured properties to infer something about the system itself and to determine unknowns.

Next we talked about the limitations that a 1-D approach imposes. For instance, if you change cross-section you really have to use a 2-D analysis. If you are going to treat two dimensions, a nice technique to use is planar Fourier acoustics. There, when you come to two dimensions, from a Fourier point of view, each of the Fourier components in the spectrum correspond to propagation of plane waves in a given angle with a given phase speed.

What else can we do? The one thing I have not really touched on is what happens if our assumption of linearity goes out the door. How does that change things? I want to wind up in the next couple of minutes by cracking the door to nonlinear acoustics and, again, I am only cracking.

[Transparency 5]

This part of the lecture is motivated by a chapter in Dr. Hamilton and Dr. Blackstock's book, Nonlinear Acoustics. Go back to the general lossless forms of the equation of continuity and conservation of momentum, as shown at the top of T50. An alternative formulation is to have the same versions of the fundamental equations in their full form and write the relationship between, say, pressure and density in the form shown in the lower right corner. It is an equation of state. Dr. Hamilton will be glad to tell you where that comes from. For our purposes λ is a property that relates to the state of the fluid. As before, we want to find the 1-D versions of these equations. However, this time we will keep terms that we neglected in the linearization process. Notice for instance, that the momentum equation contains a total time derivative, which includes both the partial with respect to time and the convective term.

[Transparency 51]

This leads us to the versions the equations of momentum and continuity shown at the top of T51. If you take the fundamental equations and apply the equation of state, you get equivalent forms shown on the third line. So what? Although it may not be obvious, these equations tell us how things propagate? By adding and subtracting these two equations, they can be recast in the form shown at the bottom of T51. J_+ and J_- are called Riemann invariants. What is important about them? Notice the form of these last equations.

[Transparency 52]

They have the same form as the 1-D version of the total time derivative, partial with respect to t, partial with respect to x. u + c and $u - c$ play the role of the speeds with which the Riemann invariants propagate. If the total time derivative of a quantity is zero, what does that mean?

Suppose a property called q that has a value of, say, 4 at one place. If that property moves with speed v, the value will not change as it propagates. In the derivation on T51, that speed is the sound speed plus or minus the particle speed. So now things do not move at the speed of sound, they move at the speed of sound plus or minus the particle speed. However, the particle speed is amplitude-dependent. So now we have introduced modifications to our fundamental equations to result amplitude-dependent effects. This brings us to the realm of nonlinear acoustics and that is where Dr. Garrett picks up after lunch.

With that, I will end. Thank you.

NONLINEAR ACOUSTICS

Steven L. Garrett
The Pennsylvania State University

DR. GARRETT: I would like to start by thanking the organizing committee, Drs. Hargrove, Bass and Atchley, for inviting me to come back to PASS again. Even though it is not clear why I should be giving this talk on nonlinear acoustics, the organizers are competent acousticians, they made the pick, and I will trust them. Logan, Hank, and Anthony were all recipients of what I used to call the "Boy Wonder Award," until Ilene Busch-Vishniac won it. It is now called the Lindsay Award (it used to be called the Biennial Award) and it is presented to a member of the Acoustical Society of America, under age 35, who "has contributed substantially, through published papers, to the advancement of theoretical or applied acoustics, or both." In principle, the organizers should know what they are doing.

When I am faced with making a presentation, I immediately ask myself two questions: The first is, "What do you want to accomplish in giving this talk?" The second question is, "How are you going to reach those objectives?

In my case, there is a third question, but I get to that on the next transparency.

[Transparency #1 – Nonlinear Acoustics]

My answers to those questions are given at the bottom of the first transparency. My motivation is to provide a general introduction to the simplest concepts and techniques of nonlinear acoustics and, in doing so, to be able to present a variety of phenomena that can be understood only by extending our treatment beyond the linear acoustic approximation. These will be phenomena that would not be observed if acoustics were strictly linear -- I want to be able to show you some of these and tell you why they are unique.

The third reason is a personal prejudice of mine. I believe that we develop a better understanding of any subject if we go beyond the limitations of that subject. If we want to understand linear acoustics, we need to go to nonlinear acoustics, because from that perspective, we really get a much better understanding of what we did when we linearized the equations which describe acoustical phenomena. What did we throw away, and what does not fit within the limitations imposed by linearity?

There are many examples of what is to be learned by "going beyond" and then "looking back." I was just having a conversation with Roger (Waxler) about radiation pressure and streaming. These are two nonlinear effects. My current understanding of those effects is quite separate, yet Roger has a very unified perspective. If I can adopt that perspective, then set viscosity equal to zero, I should get the radiation pressure back. That would provide me with a better understanding of the material.

That third objective, I think, is important. If you are going to study fluid dynamics, it is nice to study either plasma physics or superfluid hydrodynamics, where your system requires more variables than simple liquids: in plasmas you have to consider electric charge, in superfluids you have counter flow of the superfluid and normal fluid components. If you get an understanding of these multi-component fluids, you end up with a better understanding of ordinary single-component hydrodynamics. It is nice to study that, because it gives you a much deeper perspective on the simpler system and some day, a two-component fluid system might be of interest, as in the case of the liquid-vapor thermoacoustic systems now being investigated by Willy Slaton at NCPA in Mississippi.

That is what I hope to accomplish: use simple techniques, show you some new phenomena, and exploit the perspective provided by nonlinear acoustics to give you a better understanding of linear acoustics.

So, how will I attempt to accomplish those objectives? There are two ways to proceed: either the Greek Euclidian viewpoint or the Babylonian viewpoint. The Euclidian viewpoint is there are certain fundamental axioms that are taken to be true. In the case of nonlinear acoustics, those would be the full equations of hydrodynamics. Then we would use mathematics to build up the theorems and lemmas and create this beautiful structure, all based on these very clear axioms. That is the Greek approach.

There was another approach that was developed in Arabia, the Babylonian approach, which led to algebra. It is a much more egalitarian approach that says all starting points are equal. We are just going to tell you how to get from point A to point B and from point B to point C. It does not matter where we start, point C is no better as a place to start than point A, but just as there is a way to get from point A to point B, there is also a way to get to point B from point C.

I will take the Babylonian approach. There are several reasons for that choice and some of you know one some of those reasons. The fundamental reason is I am not a very competent mathematician. That stops the Euclidian approach pretty quickly. Among my colleagues, I am known as a "computational cripple." I regret that I am not good at mathematics, but like any other handicapped person, I have developed other techniques to compensate. Just as blind people rely more on their hearing than sighted people do, and deaf people use sign language.

Richard Feynman was one of the best physicists of the 20th century and quite a character. I always liked one of his observations: "One measure of our understanding is the number of different ways by which we can arrive at the same result." That is a sort of Babylonian way to reach an understanding of a particular subject. I will try to treat the same problem in several different ways that will get us to the same result. The problem with that approach is encapsulated by this Japanese proverb: "A man with one watch knows the time, a man with two is never sure."

If I build it up from a nice self-consistent, unified (Euclidian) perspective, then you have one watch and you know it works and you are always sure you know where you are going. Their is risk in the approach I will take, but I have to take the risk, because even if I could do the mathematics, I am not sure that it would be a pleasant way to spend the next two-and-a-half hours.

[Transparency #2 – Why Me?]

The third question is not a question that you have to answer, but it is a question that I obviously have to answer. After "what am I trying to accomplish" and "how am I going to do it," the question becomes, "why me?" Before I started writing these lectures, I sat down and started writing the names of -- it says "guys" here, but some of them are women -- people who are more qualified to give this lecture than I am and could do it. (I made a mistake, David Crighton passed away about a month ago, so he cannot do it, but everybody else on that list, as far as I know, is still alive and is better qualified.)

But as I said in the introduction, the organizers here are really good acousticians, so I asked myself, "What could they possibly have been thinking?" I realized there might be a good acoustical explanation: (i) We assume that the students have very little understanding because of their lack of exposure to nonlinear acoustics. (ii) We assume that the people listed by name on Transparency #2 have a really comprehensive understanding. (iii) We know from acoustics that if we form an impedance matching-later with an impedance that is the geometric mean of the knowledge, then there is going to be perfect transfer of their information to you students, as long as I represent something that is $\lambda/4$, roughly, in thickness.

[Transparency #3 – Lecture Outline]

I would like to start with an acknowledgement: What I will be presenting today is really a perspective that was generated by my thesis advisers. I learned that perspective as a graduate student, and I found it to be a very good way of understanding. I want to credit Isadore Rudnick and Seth Putterman with most of the stuff that I present correctly. Professor Rudnick is no longer with us in body, but I hope his spirit will be adequately represented. I also want to thank the Danish-American Fulbright Commission, because they gave me the luxury of living in Copenhagen while writing the notes for this lecture.

The lecture is broken into two unequal parts. Two-thirds of the lecture will be about wave-wave interactions and one-third of the lecture will be about nonlinear effects that produce non-zero time-averaged effects. What I mean by wave-wave interactions are the cumulative effects that the wave has on the medium or on some other wave. Those effects modify the way that waves propagate. I will talk about the cumulative effects of convection and nonlinear elasticity in generating shock waves; that is, I will talk about the inherent instability of a simple monochromatic wave when you include those effects in the nonlinear approximation.

Then I will use that perspective of wave-wave interactions to talk about nonlinear mixing phenomena, in particular, the parametric array. Bruce Denardo will demonstrate these effects tomorrow in the acoustic demonstration session that will be held at the Naval Postgraduate School.

I will then "shift gears" and not look at the time-harmonic effects such as harmonic generation and shock wave formation; I am going to look at the time-independent parts, forces and torques on solid bodies in intense acoustical fields. I am going to use an approach that is based on essentially the Bernoulli effect. Since the Bernoulli pressure depends on the square of the fluid velocity, there are going to be non-zero time averaged-effects, because v^2 is positive definite. When we take a time average of a positive definite quantity, we are not going to get zero.

I am going to use the Bernoulli effect to talk about acoustic levitation and acoustic levitation stability but, again, I want to give you "two watches." I will also introduce the

Ehrenfest Adiabatic Principle as another way of thinking about forces on objects placed in intense acoustic standing wave fields.

[Transparency #4 – Disclaimers and Cautions]

Now, the disclaimers: "Your mileage will vary." Nonlinear acoustics is a very rich field and, out of necessity, I had to not treat some very interesting topics. I will not cover bubble dynamics and nonlinear oscillators (I think Bruce is going to talk about that), numerical solutions and model equations like the Burgers equation. I will mention the Burgers equation only twice – that was it. (Laughter)

N-waves and sonic booms are also excluded. I am going to treat periodic disturbances, but it is possible to understand a sonic boom in the same way. I am going to talk about deterministic problems. I am not going to take a stochastic distribution of waves, traveling in a variety of different directions, and then ask about the nonlinear interactions of that ensemble. That is a problem known as wave turbulence -- very interesting but I am not going to touch it.

I am going to stick to one dimension. It is a lot harder to deal with beams. Mark (Hamilton) does with this very well. He is here at the Summer School and he may choose to say something about that problem during one of this evening's discussion sessions. I am not going to talk about streaming, which is a term that refers to steady heat and mass flows produced by oscillatory acoustical flow fields. Streaming is really at the heart of thermoacoustics. I am not going to talk about solitons, because Bruce is going to take care of that.

I also want to warn you there could be a whole bunch of errors in here. Some of them are just my errors. There are other errors that are subtler and might be more dangerous. Although Bill Gates does rule the world, you cannot run all of his software in every country. The font sets are different: I could not make any "vectors," I could not make any square brackets, and I could not do any integral signs. I used "white out" and put those symbols into the transparencies by hand. The font sets I had available in Copenhagen did not run with Mathtype™, so you have to watch those equations closely.

[Transparency #5 – Linear Acoustics]

I will start with one slide on linear acoustics. From the standpoint of this lecture the only "bonus" that you get by accepting the limitations of linearized hydrodynamics, is the principle of superposition. In linear systems you can take more than one solution and the sum of those solutions will still be a solution. We will consider two "packets" of waves shown here. The high-frequency packet is moving to the right. A lower frequency-content packet is moving to the left. In the region where they occupy the same space you will see junk; all kinds of wiggly stuff that will change with time, but they will pass through each other entirely unchanged. No new "stuff" will be generated by their interaction. When you shine two flashlight beams across each other, the light beams pass right through each other. In the linear case that is what you get: superposition is "legal."

In nonlinear acoustics, that is not true. Superposition is the first thing that is lost.

[Transparency #6 – Shallow Water Gravity Waves]

I am going to start talking about nonlinear acoustics with a problem that is not traditionally considered acoustical within the "Kinsler and Frey" perspective. My first example is a hydrodynamic surface wave. It is not a compressional wave, but we are so close to the ocean, and it is so easy to visualize surface waves, I think this is the right place to start with shallow-water waves, better known as "surf."

I am going to make that problem very simple. I am going to limit my consideration to shallow water, by which I mean that the equilibrium depth, h_o, is constant and much less than $\lambda/2\pi$. You can see automatically that my first figure is not drawn to scale; it is too deep to satisfy that last criterion.

We are going to take the simplest possible equation of state: the fluid density, ρ, equals a constant. Nothing this wave is going to do is going to change the density of this fluid, so we do not need an equation-of-state. We have an incompressible fluid and the reason it is incompressible is that if you try to squeeze it, there is a free surface and the surface will just goes up. It does not need to compress, it requires much less energy to just raise the height of surface to $h_o + h_1$.

We are also going to consider a disturbance in the mean height as a progressive wave of small amplitude, so that the modulation in the depth, which I call h_1, is small compared to the mean depth, h_o. Again, one of the reasons I start with this problem is that I can get some of my terminology in place.

The height of the free surface will be a function of position and a function of time. We are going to break the depth up into a part that is a constant, h_o, and a part that varies with the frequency ($\omega = 2\pi f$) of the disturbance, h_1 (x, t). Nonlinear effects will generate other components of the wave that may have contain other frequency components and also depend upon position or time. I am going to write that deviation from the equilibrium depth in space and time as a traveling wave going to the right, $h_1 \sin(\omega t-kx)$.

As Anthony (Atchley) pointed out in the first lecture, if I have a continuity (mass conservation) equation, a force (Euler) equation, and an equation-of-state, I can describe wave motion, in this case, on the surface of this fluid. The continuity equation is very simple. The change in height depends on the difference between the fluid that came in to a differential element of fluid of width, dx, through the imaginary surface on the left, and the fluid that exited that element through the imaginary surface on the right. If there is a gradient in the horizontal velocity of the fluid, v_x, the height is going to go up. That is exactly what is going to happen, because this wave is going to progress to the right. The fluid to the right of dx is going to start going up and then it is going to go down.

I will write all of my equations in Eulerian coordinates that are fixed to the "laboratory frame-of-reference." A lot of people who do nonlinear acoustics use Lagrangian coordinates. They like to follow a particular "piece" of gas or fluid as it moves back and forth. I am not going to do that. I get seasick. I want my coordinate system to be fixed in space and I will look at stuff coming in and going out and going up and going down, but I am not going to tag any particles. This Eulerian point-of-view is a prejudice that is consistent with my choice of experimental technique. I do not ordinarily put dye or tracer particles in my fluid. I prefer to place a pressure or velocity sensor in the fluid. My sensors are fixed in the laboratory frame-of-reference and the fluid goes by them.

The reason the surface goes up is that more fluid came in on the left than went out on the right. That is what the Continuity Equation [1] is telling me: h_1 is the change in height with time, and v_x is the x-component of the velocity, which is much larger than the y-component of the velocity.

I have the Euler Equation [2], which tells me that the acceleration of the fluid particles in this direction is $-1/\rho$ times the gradient of the pressure. In this case, the pressure is just the hydrostatic pressure, ρgh. I can take that derivative easily because ρ is a constant, since we are considering an incompressible fluid. The gravitational acceleration, g, is a constant, except in Denardo's demonstrations (he likes to modulate g) -- we are not going to do that -- and I am left with just the derivative of the height with respect to position.

[Transparency #7 – Wave Propagation Speed]

At this point I could find the speed of sound by doing just what Prof. Atchley did in his lecture. Equations [1] and [2] are two coupled linear differential equations. I could set the determinant of their coefficients equal to zero. That would give me a phase speed, $c_{grav} = \omega/k$, and the result would be $c_{grav} = (gh_o)^{1/2}$.

Or I could take the time derivative of the continuity equation, subtract h_0 (just a constant) times a spatial derivative of the Euler equation, and end up with the Wave Equation [3] for shallow water gravity waves. I am glad to see that I am not the only person in this room who contends that the wave equation is THE most useless equation in acoustics -- maybe we will talk about that in the discussion sessions. We are familiar with the Wave Equation and we recognize that that combination of constants that multiplies the second spatial derivative in the second term of [3] is the velocity squared for the propagation of waves in one direction or the other. I have written that result down in [4], but I did not have to form a wave equation to do that. Equation [4a] is a more complete expression that is true for all depth, but if you take the limit for long wavelengths and small depths, $\lambda >> h_o$, it will reduce to [4].

There are a couple of interesting results that come out of this analysis immediately. We have an expression for the sound speed [4]. That sound speed depends on the depth, h_o. We just said that this wave changes the depth, at least locally. The wave's crests make the water deeper and, therefore, the crests would travel faster than the "equilibrium" wave speed. The troughs are shallower, and in [4], h_o would be smaller, so the troughs would travel more slowly. In fact, we can take the derivative of equation [4] with respect to the depth. The change in the equilibrium sound speed with height is just one-half of the equilibrium sound speed divided by the height [5].

When we create a disturbance in this fluid, the depth changes and, therefore, the sound speed changes locally. There will also be a change in the propagation speed due to the motion of the fluid itself – a kind of self-Doppler Shift. If we assume traveling wave solutions (as we did on Transparency #6), the continuity equation [1] dictates that the ratio of the change in height, h_1, to the mean depth, h_o, is equal to the velocity in the longitudinal direction, v_x, divided by the wave propagation speed, c_{grav}. We will now define a useful dimensionless quantity called the Mach Number [6]. It is the ratio of that longitudinal fluid velocity, v_x, to the equilibrium wave speed at the equilibrium of depth, c_{grav}.

You can see that based the change in height, there is also a change in wave speed. So $\partial h/\partial v_x$ is just h_o over the wave speed [6]. We used the continuity equation to give us the freedom to express the amplitude of the wave in terms not necessarily of its height, although that is a good way to do it; we may choose to express the amplitude in terms of the horizontal velocity.

That choice turns out to be very useful.

[Transparency #8 – Cumulative Waveform Distortion]

This transparency is the single most important transparency in the first hour of this lecture, so I am going to linger with this transparency for a while. If you understand what is going in Transparency #8, and can connect it with waves out there on the beach, then everything else I am going to say about nonlinear distortion should follow by analogy. When I get through this, if these concepts do not make sense to anyone, please stop me.

To summarize what I have just said, the presence of the wave changes with wave speed in the medium. It is important to recognize at this point that I have just violated the most precious assumption of linear acoustics: the wave does not affect the medium in linear acoustics. I have just shown that the assumption is not true, the wave does affect the medium and it affects it in two ways.

First, since the wave causes the fluid to move with a strong component in the propagation direction, there is a convective velocity correction to the wave speed. There is a Doppler shift that is added to or subtracted from the wave speed. When the fluid is moving to the right and the wave is moving to the right, the wave speed is higher than its equilibrium value, c_{grav}. When the fluid is moving to the left, it reduces the wave velocity to a value that is lower than the equilibrium wave speed. The velocity also changes with depth [5], so as the surface of wave gets higher, the speed goes up and, as the surface of the wave goes down, the wave speed goes down.

I can summarize these two effects if I define a local propagation speed that is dependent on the amplitude of the wave [7]. I could have expressed the "local" speed in terms of the excess height of the wave, h_1. That might seem more natural, since we think of the amplitude of the wave in terms of the excess height, but it is much more convenient to describe the wave amplitude in terms of the longitudinal velocity, v_x. By choosing to express the wave amplitude in terms of the velocity, I can take the Doppler (convective) correction, v_x, and add it directly the equilibrium value of the wave speed, c_{grav}. The change in the local sound speed with depth can also be expressed as a function of the longitudinal velocity by combining [5] and [6]. It is the equilibrium part plus the velocity plus the change in the equilibrium velocity with height.

I end up saying that the "local velocity," the amplitude-dependent velocity, is the equilibrium velocity plus three-halves v_x. What we are saying is that the Doppler shift, due to the convective velocity, contributes one part and the change in depth contributes half a part, so the convective part is twice as important in shifting the velocity as the change in depth. That is the meaning of the three-halves at the right-hand-side of [7]. Convection and the change in wave speed with depth both affect the wave in the same way. They do not necessarily have to add up, and I will show you some deviant cases later in this lecture. In this case, for surf, they add up and in most cases they add up.

Now, we can re-examine a wave that was initially sinusoidal. If we move along with the wave, by traveling at the same speed as the "zero crossing" of the wave, that is, we move along at just the equilibrium wave speed, we find is that as time goes on, the sinusoidal waveform starts to distort. The reason should be very obvious. The crests are traveling faster than the zero crossings. The troughs are traveling slower than the zero crossings. Since we are moving along at the zero crossing speed, we see the crests advancing and, of course, in proportion, the other parts of the wave with $h_1 > 0$ are also advancing. We see the troughs retarding (and, again, in proportion, the other parts with $h_1 < 0$ are retarding). That is forcing that waveform to bend over.

As we go farther and farther, the crests are going to get farther and farther ahead, so these effects are cumulative. They do not average out to zero, even though the heights average out to zero. They are cumulative effects.

We can ask ourselves when this simple picture starts to fall apart. The simplest answer is not the conventional answer. We can ask how far does this wave have to go in order for the crest to actually get ahead of the zero crossing and the trough to get behind the zero crossing? In a system that does not have a free surface, the wave would become multiple-valued – one position could have three different values of h_1. Of course, out on the beach it does become multiple valued. Waves curl, they do that. Then they crash.

In other systems that do not have a free surface you cannot get a multi-valued result for the amplitude of the wave at a single location. At the point the wave front is vertical, there will be an infinite gradient. If there are any dissipative effects whatsoever, that large gradient is going to prohibit a vertical wave front.

At this point we are in a position to determine how far the wave would have to propagate before the gradient would become infinite. Another way to pose the question would be to ask ourselves what would be a characteristic length or characteristic distance that a wave had to travel before it became a shock wave? It would be easy to calculate that distance because we know that the crest of the wave is traveling at a speed that is three-halves v_x greater than the zero crossing. For the crest to catch-up to the zero crossing, it would have to advance by a distance that is a quarter-wavelength, $\lambda/4$. The distance the wave would travel in that time would be the equilibrium wave speed times the time it would take to for the crest to advance by a quarter-wavelength.

What is that time? It is $\lambda/4$ divided by the additional speed, $3v_x/2$. So this is what we might call the shock-inception distance, except that you do not want to use $\lambda/4$. It turns out that the waveform becomes vertical at this the zero-crossing when the crest has advanced by $\lambda/2\pi$. But you get the basic idea whether you use $\lambda/4$ or $\lambda/2\pi$: We want to know how far the wave has to go before you reach an essential discontinuity. That is known as the discontinuity distance or the shock-inception distance, and that is what I have calculated in [8].

So let's review this because Transparency #8 is the most important slide of this hour. The amplitude of the wave affects the local propagation speed. It affects it in two ways. It adds a Doppler shift, a convective term, due to the longitudinal velocity of the fluid. It also changes the depth, so it changes the wave speed. Those effects cause the crests, which have the positive velocities to speed up, the troughs to slow down. The waveform distorts. There is a characteristic distance, D_s, associated with this distortion process [8]. That distance, in units of the wavelength, for shallow water gravity waves, is the wavelength divided by 3π times the Mach Number. As you consider waves of larger amplitude, the shock happens at a distance that is closer to the source.

Who is uncomfortable so far?

[No response]

Then I am not trying hard enough. Let's move on.

[Transparency #9 – The Grüneisen Parameter]

We are not generally, as acousticians, interested in surf -- at least in the professional sense. Typically, we are interested in the nonlinear behavior of gases or gas mixtures, like air, or water and other fluids, or solids. As long as you understand the basic concept illustrated in Transparency #8, those ideas can be applied to other continuous media. The $3v_x/2$ term was a characteristic of the shallow water wave problem. Other systems will have different coefficients, depending upon the nonlinear characteristics of the media, but otherwise the distortion processes will be identical.

To treat the more general case, I will follow the same strategy. I will let the local sound speed depend on the longitudinal velocity in the medium that is generated by the presence of the wave. I will introduce a new parameter, capital gamma, Γ, which will provide the factor that multiplies the convective velocity to incorporate the change in the equilibrium value of the wave speed due to the presence of the wave. I use Γ because my graduate education was in the area of condensed-matter physics. In that field, a Grüneisen constant was introduced to account for the elastic nonlinearity that produced the thermal expansion of solids.

Nobody but Seth Putterman and me uses that designation, but it is not important. Some people call the parameter of nonlinearity β. A common choice in acoustics is to specify a B/A coefficient. I will show you how to make connections to all those, but I am going to use Γ because it makes it easy for me to track the strength of the nonlinear effects.

All I need to do is find the value of Γ for a particular situation and I will be able to take everything I just explained to you about surf and apply it to some other medium. For instance, if the speed of sound depended on density, then Γ would be one plus the change in sound speed with density times the change in density with velocity as shown in [10]. Of course, that is easy to evaluate because the continuity equation [11] relates changes in density and changes in velocity. I chose that particular example because, later, when we talk about parametric arrays, [10] will arise directly out of the hydrodynamic equations.

Let's apply this approach to sound in air or sound in any other ordinary gas. We need to determine Γ_{gas}. Then we throw that Γ_{gas} back into the "surf equations" and we are home free.

[Transparency #10 – The Grüneisen Parameter]

The speed of sound in an ideal gas [12] is given by $c_o^2 = \gamma RT/M$. In this case, lower-case gamma, γ, is the ratio of the specific heat at constant pressure to the specific heat at constant volume. R = 8.31451 J/mole-°K, is the universal gas constant. T is the absolute (Kelvin) temperature, and M is the molecular weight (in Kg/mole) of the gaseous medium.

We can differentiate [12], but before we do, recognize that the speed of sound in an ideal gas is controlled by only one parameter. Since the molecular weight and the ratio of specific heats are not affected by the presence of the sound wave, the temperature controls the sound speed. The pressure change is cancelled by the density change. The way a wave changes the sound speed in a gas is by changing its temperature. The relative change in sound speed is one-half the relative change in absolute temperature [12].

Now we build the chain that connects the variation in sound speed with the variation in the oscillating longitudinal fluid velocity produced by the sound wave. The local propagation speed, as a function of the velocity amplitude, is the equilibrium sound speed, plus the convective contribution, plus the change in sound speed with temperature. We can relate to the change in temperature, T_1, to the change in pressure, p_1, through the adiabatic gas law [14, 15]. We then relate to the change in pressure to fluid velocity through the Euler equation [16]. We put it all together, holding the entropy constant (since sound propagation in gases is adiabatic), so all we have are a few derivatives to evaluate.

We have already calculated $(\partial c/\partial T)_s$ [12]. $(\partial T/\partial p)_s$ can be obtained from the adiabatic equation-of-state. This beautiful trick in [14] was shown to me by Tom Gabrielson. I used to have a harder way of doing this. We know that pV^γ is a constant and we know the ideal gas law: $pV=nRT$. Again, in this case [15], I prefer to take the logarithm before I differentiate, when I have expressions that involve power-law behavior.

As before, we can use the hydrodynamic equations, but now we want to relate pressure to velocity. That sends us directly to the Euler equation [16] and we find that the changes in pressure are related to changes in velocity by the specific acoustic impedance, $\rho_o c_o$. We substitute values of all three of those derivatives into [13] and we find [16] that the Grüneisen coefficient, Γ_{gas}. For a gas this is 1 (the convective part) plus $(\gamma-1)/2$ (the nonlinear equation-of-state part), which I can simplify to $(\gamma+1)/2$.

For an inert gas, γ is five-thirds, so the distortion parameter is $\Gamma = 1.333$. The fractional part is due to thermodynamics; the integer part is due to convection. For air, the ratio of specific heats is seven-fifths, so you get the convective part plus another 20% due to the thermodynamic part.

The reason I point this out is that some people tend to believe that the nonlinearity of the medium is an important component. It is certainly important, but it is not nearly as important as the convective part in many cases. You could have a perfectly linear medium, say a gas that compresses isothermally, and you will still get distortion. You would get only 20% less nonlinear distortion than if you sending large-amplitude waves into air. The nonlinearity is built into the hydrodynamics; it is not strictly a property of the nonlinearity of a specific medium.

[Transparency # 11 – Shocks in Air]

Let's apply this result in order to get a feeling for what we are talking about in terms of shock distances and distortion. We start by using the loudest sound wave that we would normally encounter in ordinary air-borne acoustics, 120 dB_{SPL}. That is the threshold of pain. That corresponds to a 20 Pa_{rms} disturbance, so peak pressure amplitude, p_1, the amplitude of a peak or trough is 28 Pa. The Mach Number, which is the critical nondimensional measure of how strong that wave is, is given by v/c or $p_1/\gamma p_o$. That amplitude corresponds to a Mach Number, which is only 200-ppm [18]. The loudest sound (peak pressure) you can hear without really hurting yourself is only 280 ppm of the mean atmospheric pressure, less than 0.03%.

We can plug that result for the Mach Number, M, back into the expression that I had for the shock inception distance [8]. We now must use the appropriate value of Γ for air, $\Gamma_{air} = 6/5$, instead of $\Gamma = 3/2$ that was used for shallow-water waves. We find that the wave will form a discontinuity, if it is one-dimensional, in air at a frequency of one kilohertz, after traveling 230 m away from its source [19].

Does some other effect take over before you reach this the shock formation distance, D_s, to stop the shock-wave formation? Is it possible that just simple thermoviscous dissipation will attenuate that wave and it will not maintain that original Mach Number out to D_s? If you look at Prof. Bass's work on attenuation in the atmosphere, you find that at 1 kHz in dry air, the distance that it takes for the wave's amplitude to decay to 1/e of its initial value, that is, the exponential attenuation length is 4.3 km. This suggests that shock wave effects will be more important, in this case, than classical attenuation effects.

We form another dimensionless quantity called the Goldberg Number to compare the nonlinear effects with ordinary (linear) dissipation. The Goldberg Number is the ratio of the distance that it takes for the wave to attenuate by 1/e, divided by the distance that it takes to form a shock, D_s [20]. When the Goldberg Number is bigger than one, shock waves will form; hence this wave will shock before it attenuates.

If we increase the frequency by a factor of 10, if we go up to 10 kHz, then the wave would shock in 1/10 the distance if the attenuation was the same at 10 kHz as it was at 1 kHz. The shock formation distance is scaled by the wavelength. If I raise the frequency by a factor of 10, I shorten the wavelength by a factor of 10. That 10 kHz wave would shock in 23 meters, except that the attenuation increases with the square of the frequency. The attenuation length drops from 4.3 km to only 43 meters.

In fact, the Goldberg number goes from our initial (1 kHz) case of 19, which is clearly shock-controlled, to a lousy factor-of-two at 10 kHz. This indicates that the 10 kHz wave is going to shock, but it is just barely going to produce a slight "discontinuity" for a little while and will then attenuate away in a more linear fashion. When we get up to 100 kHz, a very high frequency, where the shocking distance would be only 2 m, it will never form a shock front, because the attenuation has gone up so high that the wave will attenuate before the crests catch up with the troughs.

There is a whole range of these Goldberg Numbers and for the case of deep ocean waves Goldberg Numbers are always about a million. Viscosity is entirely irrelevant for deep ocean surface waves. All of the interactions of the waves on the surface of the ocean are completely dominated by nonlinearity. Navier can take off with Stokes and they can have a good time. You just do not need them for that case. The Goldberg Number will tell you whether you are going to be controlled by nonlinear distortion or irreversible thermal conduction and viscous losses that Professor Atchley talked about earlier.

[Transparency #12 – Stable Waveforms]

If you are in the large Goldberg Number regime, then you are going to form a shock front. Every monochromatic periodic wave that you excite, for a sufficiently large Goldberg Number, will end up looking like a sawtooth wave. If you put in a sine wave you will get a sawtooth. If you put in a sawtooth wave, you will get a sawtooth. Put in a triangle wave, you will get a sawtooth. Put in a backward sawtooth, you will get a forward sawtooth.

It is interesting, if we are going to study nonlinear acoustics, to understand what goes on in a sawtooth wave. We can represent a sawtooth waveform as the superposition, in the Fourier sense, of sinusoidal waves of progressively higher frequency that are harmonics of the fundamental [21].

We can analyze the dissipation created by the large gradients produced by the shock front in at least two ways. One way involves writing conservation equations for the pressure and the density behind and in front of the shock. Those conservation equations [22, 23, and 24] are known as the Rankine-Hugoniot relations. From there, one can show that across the shock there is a discontinuity in the entropy. The dissipation that takes place at the shock front is third-order in the relative pressure difference ahead and behind the shock [25]. That approach is explained nicely in Landau and Lifshitz, *Fluid Mechanics*.

[Transparency #13 – Shock Attenuation]

If we are in the regime where the wave dissipation is due entirely to the shock front; the fully-developed shock wave limit, then there are large gradients. As shown in the figure, a straight line that pivots about the zero crossing can represent the non-shock part of the waveform. If we are moving in a frame that is moving along at the same speed as the zero crossing, each point on the waveform which is not part of the shock front is going to more forward by an amount that is proportional to its amplitude. That corresponds to a solid-body rotation.

In fact, the wave wants to pivot about this zero-crossing point. To do that, it is going to have to eliminate the part of the wave represented by the smaller shaded triangle as it passes through that shock front. I have reproduced this geometrical construction from my notes on a course given by Prof. Rudnick at UCLA back in 1977, but I could not find this elegant geometrical approach presented anywhere in the scientific literature.

As the line rotates, then its intersection with the vertical shock front will drop. I have exaggerated the infinitesimal drop, du, for visual effect. The part of the construction that moves ahead of the shock travels with the excess velocity in the moving frame and advances by a distance Γudt, in a time, dt. The two shaded triangles are similar, so the ratio of the opposite to the adjacent legs [26] is equal; hence we have an expression for du that can be integrated to produce [27].

This is an interesting expression for attenuation. It is not what we are used to; it does not lead to exponential decay. It is an algebraic attenuation that has a very long time tail, which you do not observe, since at long times, the linear thermoviscous dissipation takes over. What you find is that the shock will attenuate much more quickly than classical exponential decay. What is initially even more surprising is that the decay rate does not depend on magnitude of the transport coefficients. When you talk about attenuation in linear acoustics, you expect to see loss rates that are proportional to viscosity and thermal conductivity. The only "material" parameter you see in [27] is Γ. I would like to explore what that means.

[Transparency # 14 – Shock Front Thickness]

Rudnick's geometrical construction says that some amount of energy is going to have to be removed from the system to allow the sawtooth waveform to evolve by the "solid body rotation" model. To do that, the gradient is just going to get as steep as it needs to be. If the wave amplitude is larger, then the gradient is going to get steeper, so that the steeper gradient can dissipate the required amount of energy. If it is not such a large sawtooth wave, then that gradient will reduce as necessary, but the energy loss does not depend on the dissipative coefficients directly; those dissipative coefficients will determine the magnitude of the gradients across the shock front.

I am going to use that picture to calculate the thickness of that shock front. This is another interesting problem in nonlinear acoustics. As always, I am going to cheat, since the correct answer is already known by a more careful analysis. We know that the kinetic energy contained in one wavelength, E, is proportional to the velocity squared [29]. That proportionality constant, k, could be ½ρ. The length of a unit cross-section of one wave is λ.

Since E is proportional u^2, the change in energy divided the energy is two times the change in velocity divided the velocity -- that is just a differentiation of [29]. Our construction happens, conveniently, to have produced [26]; just the quantity we need, du/u. If I plug du/u into [30], I obtain the relative change in energy, dE/E, in terms of the relevant parameters of the problem, M, Γ, and λ.

On the other hand, if we go to classical acoustics and we ask what is the thermoviscous attenuation, we can write [28], which is probably familiar to most of you. The attenuation rate in space, α, is some product that depends on, ω^2, the frequency squared, μ, the shear viscosity, a relaxational component, ξ, if you have a media other than an inert gas, κ, the thermal conductivity, c_p and c_v, the specific heats at constant pressure and constant volume. What may be less familiar is my preference for gathering these details describing the transport properties into a single "collision" time, τ. Using that time, and the equilibrium sound speed, c_o, I can define a characteristic length that is related to the mean-free-path, *mfp*, between collisions of the molecules which make up a gas.

I am going to exploit the fact that the frequency squared determines the attenuation. That way I can approximate the thickness of the shock front is by ignoring all but the highest frequency component (remember the Fourier expansion of the sawtooth in [21]), since that will be the frequency component that attenuates the most amount of energy.

The logic behind this approach is based on my assumption that the thickness of the shock front, $\delta = \lambda_{max}/2$, is equal to one-half wavelength of the highest frequency component, f_{max}, required to represent the waveform in the Fourier superposition [21]. Using [28] as the expression for the thermoviscous loss, the relative energy loss due to thermoviscous attenuation of that single high-frequency component, $dE_{T\text{-}V}/E_{T\text{-}V} = -2\alpha\ dx$. Equating that result to the relative energy loss due to the shock front [30], we obtain an expression for the shock thickness [31].

That approximation to the thickness turns out to be $\pi^2/2\Gamma$ times the mean-free-path. This is not exactly right, because I was fudging here by throwing away all the lower frequency components, but I end up with a result for the shock thickness that is related to the mean-free-path divided by the Mach number. The stronger the shock, and the larger the value of M, the thinner the shock front will be. This is essentially a quantification of the argument presented previously that explained the absence of the transport coefficients from the shock attenuation result [27].

For the example I gave you, which is about the weakest shock you can think of, a Goldberg Number of 19, that thickness is only 70 microns. It is 20,000 mean-free-paths, but it is very small compared to the wavelength; that is, the wavelength divided by the shock-front thickness, even for a very weak shock, is a factor of 5000. That shock thickness is 1/5000 of the wavelength. If you upgrade to a really good shock, with $M \approx 0.1$, then the thickness gets down to be on the order of only about 40 mean-free-paths; less than 0.2 microns.

The picture that I want you to take from this discussion is that once you are fully developed as a shock wave, that the back of the sawtooth wave joining the successive fronts is going to rotate as the wave progress. That energy loss is going to be absorbed by the steep gradient represented by the shock front. The gradient is going to adjust its steepness so that it can get rid of the part of the "triangle" that does not belong there, since it would make the front multiple-valued. If we accept that it is a thermoviscous process, although that is irrelevant, we can say what the shock-wave thickness will be due to thermoviscous processes.

MR. TUTTLE: Can you show us on that graph where the thickness is?

DR. GARRETT: At the scale of the construction in Transparency #13, a vertical line represents the shock front. In a weaker shock it will be tilted slightly. Remember, even in the weak shock limit, the "tilt" is one part in 5000. It is virtually vertical and that is why I said that you have to really go down to the molecular level to "observe" the tilt. The tilt is on the molecular level and the wavelengths are on the hydrodynamic level. That was the outcome of the previous development.

[Transparency #15 – Distortion in Liquids]

It is very common to extend this treatment of shocks to liquids. A lot of work is done, particularly in biological liquids. The people who work in that area prefer to talk about the B/A coefficient. In their literature, you will see B/A's for brain tissue, for bovine serum albumin, and other stuff that I lump into the same class as hamburger meat. The researchers who work with liquids and biological materials choose to express the nonlinearity in the equation-of-state for their favorite media by expand it in terms of a power series in the density of the media. That power series is known as a virial expansion [32].

The first term is proportional to the deviation of density [32a]. The second term is the coefficient of deviation in density squared [32b]. The third term is the coefficient of the relative density deviation cubed [32c]. That approach to the equation-of-state is called a virial expansion. Van der Waals has another way of writing an equation-of-state. There are lots of ways of writing it. The adiabatic in gas law is another equation-of-state.

I have taken those various virial coefficients and expressed them in terms of thermodynamic derivatives [32a-c]. Those results express the virial coefficients in terms of the change in sound speed with pressure. A practical choice, since measurement of the change in sound speed with pressure is a pretty common way to determine the nonlinearity in the equation-of-state. We can express the sound speed in terms of those virial coefficients as well [33]. That form is useful to me, because I said there is the convective contribution and the sound speed contribution to the nonlinear distortion process.

The lowest-order correction is B/A times the relative deviation in density. If we take the expansion to the next order, it adds a further correction, which is C/2A times the relative density deviation squared. If we stop the expansion here, we can identify those terms. Γ is just 1+B/2A. You see that result in almost every nonlinear acoustics book (except they will not call it Γ). The distortion term is 1+(B/2A), the "one" being the convective contribution.

I am going to take you a little further, because there is something interesting that happens if we go to the third term. Again, as I say, if you step beyond what you are setting out to understand, you will have a more complete picture when you look back. I hope the following will give you a broader perspective.

The third term is positive definite, so if the third coefficient in [33], C/2A, is negative, then when the density increases, the third term makes the speed decrease. When the density goes down, the speed still decreases, because $(\delta\rho/\rho)^2$ is positive definite. If the C/A term is important, then instead of the convective term and the sound speed term having the same sign, the terms will have an opposite effect for half of the acoustic cycle.

[Transparency #16 – Double Shock Formation]

For half of the cycle, both contributions will add up and distort the wave. For the other half of the cycle, they will cancel each other. The completeness of the cancellation will depend on the amplitude of the wave, as well as the relative sizes of A, B, and C. That is what I have shown in the figure. If you include the third term in the virial expansion, then there is a critical amplitude, v_d, corresponding to the complete cancellation between the B/A term and the C/A term. When the amplitude is much less than that critical amplitude (in fact, in the top waveform, just half of the critical amplitude), then the crests travel faster than the troughs and the troughs travel slower than the crests. This produces a waveform that is similar to the shock-wave formation waveform on Transparency #8. You can see there is a little bit of asymmetry between crests and troughs, but it is not dominant.

When you get to an amplitude where the peak particle velocity is equal to this critical velocity, then you will notice that the crests did not move relative to the zero-crossing at all. This is because the B/A term is speeding it up by exactly the same amount, at that amplitude, that the C/A term is slowing it down. The crest at this very point stays put, of course, but the trough is shocking like crazy.

If you go to larger amplitude, that C term takes precedence during more of the cycle. At small amplitudes, the B term is still having its way, but when you start getting to the larger amplitude portions of the wave you can observe the "retardation" during both half-cycles. When you get to very large shock amplitudes, you produce two shock fronts per wave.

DR. BASS: You made the assumption that B/A is a composite quantity. Why did you make that?

DR. GARRETT: Because I have not gotten to my next transparency. Prof. Bass is exactly right. In ideal gases, B/A is a positive quantity; in most substances it is a positive quantity. In liquids near their critical points, it is necessarily positive. In solids it may not true. In third sound on thin films of superfluid helium, it is not always true. The sign of B/A depends upon film thickness. Second sound, which is a thermal wave in superfluid helium, it is not always positive. In the case of second sound, it is a function of temperature. I am just working you up to the deviant cases because I live for the deviant cases.

You can see if you continue this, you will not get the crests always traveling faster and the troughs always traveling slower, if there is a higher-order correction in this picture.

[Transparency #17 – Double Shock Occurrence]

The question is do we ever see double shock formation? In ideal gases C/A is, in fact, negative. The ratio of specific heats, γ, is always less than or equal to 5/3, so by [36], the C/A coefficient is negative. But if you plug that result into the expression for the critical value of the Mach Number [35], the denominator vanishes. You cannot produce double shocks in an ideal gas no matter how strong the shock wave.

Classical liquids: You can do the same thing for a classical liquid. Both the C/A coefficient and B/A coefficient have been measured for water. You need to get to velocities that are 20% higher than the sound speed. You need supersonic amplitudes. It is not going to happen since the required pressure amplitude is 26,000 atm, well past the cavitation threshold for water. What is the record for water, Tom?

DR. MATULA: Two hundred atmospheres, negative.

DR. GARRETT: So you are two orders-of-magnitude away from ever seeing this shock doubling in water.

It turns out that in liquids near their critical points this effect has been observed. One of the big mistakes I made in my career was writing a theoretical article on this effect and publishing it in the *Journal of the Acoustical Society of America.* Unfortunately, any time the editors of JASA saw weird shock-wave calculations from really good mathematicians, they would send those articles to me for review. After hearing enough lame excuses, they finally realized I could not referee these articles. There are several such articles published in the *Journal of Fluid Mechanics* and in JASA, where you do see this double shock behavior in liquids that are close to their critical points.

You should be able to see this effect in superfluids, because the fluid velocity depresses the superfluid fraction and slows down the propagation speed [37]. It does not matter whether the velocity is in one direction or in another direction. It has been observed and there are places where you can get these deviant forms of shock waves.

DR. ATCHLEY: Has anybody looked at sonoluminescence?

DR. MATULA: Not for water, not that I know of.

[Transparency #18 – Reverse Shocks]

DR. GARRETT: I would like to answer Prof. Bass's question, which was based on the fact that there are circumstances where the equation-of-state part has a different sign than the convective part. One of the cases that is closest to what we talked about when we started this presentation with surf is what is known as third sound or waves in thin films of liquid helium on a nice flat substrate.

There the dominant restoring force for these very thin films is not gravity; it is van der Waals force. I just heard that it is also van der Waals force that keeps geckos on vertical walls and ceilings. It was just published in an article in *Nature*. Anyway, the van der Waals effect is what is pulling the free surface of the helium film down toward the solid substrate, not gravity, because we are at really talking about films with thickness that are comparable to atomic dimensions. That van der Waals force is proportional to the inverse-fourth-power of the thickness, due to fluctuations in the electromagnetics. The fluctuating dipoles are attracted to their "images" in the substrate, so you have a fourth-order dependence of force on the thickness.

Since the wave propagation speed is determined by the acceleration of gravity and the depth in a shallow-water wave [4], the force that appears in the wave speed, is now van der Waals attraction, instead of gravity [38]. You find out that in thin films, the thin parts travel faster because the restoring force is greater. For a sufficiently small mean depth, when the film gets thinner, the wave speed goes up; just the opposite of the result [4] derived for shallow water gravity waves.

In third sound, there is a thickness where the convective part is exactly canceled. For even thinner films, the waves break backward. For thicker films they break forward, like ocean waves. If you are an electron, you can surf on a thin sea of superfluid helium and never worry about being wiped out. The waves break behind you.

This also occurs in another superfluid helium sound mode. There is a thermal wave called second sound. Its speed will decrease with increasing temperature. This is just the opposite of an ideal gas, where the speed increases with increasing temperature. Therefore, when a second sound "crest" becomes hot, it slows down. It can slow down enough to compensate for the convective part. If you are at a temperature close to superfluid transition temperature, the lambda temperature, T_λ, the second sound wave will form a trailing-edge shock front. If you are at a temperature well below the T_λ, it will form the shock at the leading edge. In between, there is a temperature, 1.884 °K, where it will not break at all [S. Putterman and S. Garrett, "Resonant Mode Conversion and Other Second-Order Effects in Superfluid Helium," J. Low Temp. Phys. **27**(3/4), 543-559 (1977)]. The nonlinear effects will have to appear at higher order.

The same is true, as Phil pointed out to me once, in shock tubes. They are solids.

DR. MARSTON: You can have anomalous waves.

DR. GARRETT: They use those in shock tubes. When they want to make a high-amplitude pulse with a sinusoidal shape, they pass it through some material that has shock waves break backward. If they make the material thick enough, the shock can "un-distort" and it returns to the desired sinusoidal shape.

DR. HAMILTON: Glasses will do that.

DR. GARRETT: Thank you.

[Transparency #19 – 1st Review and Summary]

This concludes the first portion of my lecture. Let me go over what I think I have tried to present:

Self-interaction: The presence of a wave affects the sound speed. It does it in two ways. It does it because there is a convective contribution to the sound speed due to the fluid particle velocity in the direction of propagation. There is also a constitutive contribution that comes out of nonlinearities in the equilibrium equation-of-state.

Nonlinear effects accumulate with distance: The nonlinear effects do not change sign, so they do not cancel; crests always advance and troughs retard. These effects accumulate with distance. If nonlinear effects dominate dissipative effects, you always end up with sawtooth waveforms. The fate of those sawtooth waveforms is that they will attenuate faster than you would expect based on the linear dissipation mechanisms due to thermoviscous effects. They attenuate in an algebraic way, not in an exponential way, until the algebraic long-time tail intersects the linear-acoustic exponential decay. That is basically the waveform "life cycle."

Fully developed shocks have sawtooth waveforms: If an initially sinusoidal wave has a Goldberg Number greater than unity, it forms a shock beyond D_s. The steep gradients in the shock front dissipate energy until the wave's amplitude is small enough that the linear-acoustic thermoviscous effects now provide the primary attenuation mechanism.

There are interesting deviant cases: There are places where the convective and the constitutive parts have opposite sign. There are higher order contributions, say, the C/A term, which may not change sign, while other parts are changing sign, thus leading to other types of shock formations.

With that, you have 10 minutes for a break. When you return, we will move on to looking at this whole process again, but from a different perspective.

DR. GARRETT: Ready for part two of the Babylonian captivity?

As I promised, we are going to look at the same distortion phenomenon again, but we are going to look at it in a slightly different way. I am going to attempt a more formal approach. It is still based on the same concepts we introduced in the first hour, using convective and constitutive nonlinearity. I am going to be a little more careful about breaking the acoustic fields into, say, a first-order contribution and a second-order contribution. This approach was first executed in a very elegant way, at a time when the Americans were shooting each other in the Civil War, by a man named Earnshaw (I do not know his first name) in England.

[Transparency #20 – Waveform Instability]

Earnshaw looked at this problem by saying just what we said during the first part of this lecture. We take a look at an ordinary first-order linear sound wave with some amplitude, v', and we write the original sound wave as $v_1(x, t) = v' \cos(\omega t - kx)$. He factored the angular frequency, ω, out of the phase factor as shown in [39]. At that point, he could have said, "All right, I'm not going to restrict the velocity be a constant, c_o, but I'm going to substitute this local velocity into this equation [40] to generate a correction to the original wave form [39]." A binomial expansion for small Mach Number, v_1/c_o, will generate the next order correction [41]. We have just executed an iterative process. We have substituted the simple result, v_1, into a term that will produce a correction of higher order and smaller magnitude.

[Transparency # 21 – Earnshaw Continued]

We know how to treat a trigonometric function that has an argument that is expressed as a sum [41]. We know that cos ω(a+b) is cos ωa times cos ωb minus sin ωa times sin ωb. Since cosine of a small number is one, and sine of a small number is just that small number, we obtain [42]. The first term in that expression just gives us back v_1, so it is easy to identify the second term as v_2. Substituting our expression [39] for v_1 in the second term produces an expression for v_2 with amplitude proportional to v'^2 [43].

We know that 2 (sin a) (cos a) is just sine of 2a, so we can rewrite this trigonometric product as [44]. We have this combination of factors: x, ω, Γ, times the amplitude squared divided by the equilibrium sound speed squared, times a wave with twice the frequency and half the wavelength. By doing this procedure, Earnshaw was able to generate an expression for the second harmonic component that grows linearly with distance, x, from the source. When this pure sinusoidal wave starts to propagate away, the second harmonic gets bigger and bigger linearly with distance. I actually like to write [44] in a slightly different form on the right-hand-side of [44]. You end up with a second harmonic propagating component that is proportional to the distortion amplitude, which amplitude grows with the distance, x, scaled by the original wavelength.

We find that the first-order solution is unchanged, but that is an artifact of our approximation. You cannot be generating energy at 2f and not taking it out of somewhere. Obviously, it must be coming out of the energy at frequency, f. But because we are doing this in a kind of sloppy iterative way, which I will correct in a moment, what we do is we just recover the fundamental and we get a linearly growing second harmonic. If the magnitude of the v_2 term is small, then the decrease in the v_1 term should be negligible.

We can ask ourselves where this approximation should become invalid. Since we were assuming that each successive contribution was smaller than the previous one, $v_2 \ll v_1$, we can certainly say that that the approximation should fail when those two contributions are of equal size, $v_2 \approx v_1$. If we ask ourselves at what distance does the v_2 term equal the v_1 term, so we know we are in trouble, since our assumption is being violated, what do we get?

Surprise! You get the twice the shock-inception distance. At this crude level, this all makes sense. If we use this iterative solution to generate the weak distortion product, it will work quite well. We have to account for the energy lost in the fundamental, but we will do that.

This is a nice way to look at that shock formation, I believe. We are just taking the velocity and substituting it back into the original wave to modify the local sound speed. Even in this case, the trigonometry is a bit messy, so it is not something that you want to do very repeatedly to generate the higher frequency components required to "fill out" the Fourier representation of the sawtooth wave form [21]. Let's say we wanted to get the third harmonic, and the fourth harmonic, *et cetera.*

[Transparency #22 – Higher Harmonic Generation]

There is actually an easier way that a young Dr. Hargrove came up with in 1960. He reverted to that geometrical description again that says that the excess velocity is proportional to the waveform's velocity amplitude at each position. We know that the crests are traveling faster than the troughs, so we can parameterize the distorted waveform. We can say that the y-amplitude equals sin θ, but the value of that amplitude is going to occur at a distance farther along the wave that is proportional to that amplitude [47]. The obvious scale length for the parameterization of the x-position is the shock inception distance, D_s, as shown in [46].

We are going to plot this parametrically. We are going to plot y and then we are going to use y to determine where we plot the x-value for that given value of y. You get something that is distorted, as shown there, but that we can calculate the Fourier coefficients in the usual way.

[Transparency #23 – Higher Harmonic Generation (Continued)]

In principle, we should be able to get back Earnshaw's solution and pick up the decay in the fundamental, as well as the growth of the second, third, and fourth harmonic; as high as you want to go. All of this works out because you are taking the trig function of something that is already a trig function; that process generates Bessel functions. By doing a Fourier analysis of that shape [48], you end up with coefficients for each frequency component. The individual coefficients are proportional to Bessel functions of the same integer order as each of the harmonics, divided by the argument of the Bessel function [49]. The fundamental is proportional to a J_1 Bessel function, divided by the argument of the function.

When you first look at this, you go, oh, yes, it looks like Bessel functions, because J_0 has a value of 1 and it decays with increasing argument and looks just like that. No, that is not J_o. That is J_1. For small values of the argument, J_1 actually increases linearly with argument, initially, divided by the argument that also is increasing linearly. You get something that looks fairly constant for a while (as with the Earnshaw result) but then falls off. That is where your fundamental power is going, to feed the hunger of these other growing children; the higher harmonics.

Of course, any such analysis of the growing harmonic distortion will only be valid up to the point where the gradient, that is, the waveform, becomes vertical. These calculations of harmonic distortion were done by various horrendous mathematical techniques, but Hargrove's was certainly the slickest and easiest way to obtain the desired result. What is that result?

The second harmonic grows linearly with distance. This is the distance it has traveled scaled by the shock formation distance. This second harmonic grows linearly, which is what was produced by the Earnshaw solution. I just did it for you in a completely different way. The third harmonic grows as the square of the distance from the source. The fourth harmonic grows as a cube initially.

[Transparency #24 - Kongensvej]

I will not be able to do too much detail in the next section, but what I would like to attempt is at least an introduction to the concept of a formal perturbation solution to the hydrodynamic equations. Prof. Atchley showed the hydrodynamic equations. The title of this transparency is proof that I was actually in Denmark when I prepared these notes. Does anybody know what *Kongensvej* means in Danish?

PARTICIPANT: King's Road.

DR. GARRETT: Right.

In principle, everything I have done in the sort of Babylonian way should be recoverable in the Euclidian limit. I can start by saying that we are talking here about fluids and that the thermodynamics of a single component, homogeneous, isotropic fluid can be described by two variables. Two variables form a complete description: a mechanical variable like density or pressure, and a thermal variable like entropy and temperature.

If we want to describe not only the thermodynamics, but also the hydrodynamics, we have to add three more variables; that is, we have to give every point in space a velocity. So five variables completely describe the fluid's hydrodynamic behavior and, therefore, I have to come up with five independent equations and the algebraic system then becomes "closed."

I have five equations. I have conservation of mass [50], conservation of entropy for the two thermodynamic terms [51], and conservation of momentum for the three components of velocity [52]. Again, I apologize, but the Danish font set would not support vectors in my computer. It should not introduce too much confusion. You know which variables are vectors and which are not by how they are being treated.

I should be able to regenerate everything I have done so far by taking the hydrodynamic equations up to second order.

[Transparency #25 – Formal Perturbation Expansion]

You are all familiar with the linearization of the hydrodynamics that produces ordinary linear acoustics. The equation of state has to be introduced, because you will notice there are five equations [50, 51, and 52], except they happen to contain six unknowns. They require both pressure and density, so I have to add an auxiliary equation, the equation-of-state, to relate the pressure and the density. To obtain the linear results, I have to expand the equation-of-state only to first order, neglecting the final term in [56].

[Transparency #26 – First-Order (Linear) Solution]

I end up, again, with the infamous wave equation [61], which is important in this discussion only to the extent that it is homogeneous, which means that if I ever see this combination of terms I can cross them out; I can throw them away. That combination is equal to zero; that part is very useful. The other thing, of course, is it assures us that the first-order solutions are traveling waves for velocity, density, or pressure, whatever you choose [62].

[Transparency #28 – Nonlinear Wave Equation]

Instead of stopping at linear terms, I now include all terms up to second order in the continuity equation [68], the Euler equation [69] and the equation-of-state [70]. If we examine conservation of mass [68], the derivative of density with respect to time contributes two terms. There is a first-order term, $\partial\rho_1/\partial t$, and a second-order term, $\partial\rho_2/\partial t$. There is the dot product of the velocity vector and the gradient of the density, which is a second-order term, but there is also the divergence of the velocity times the density that has both a first-order and second-order contributions. Then there is, again, the term, like the first-order term, that involves a second-order quantity, which makes it also second order.

These are all of the terms that you must keep, a lot more than in the linear case. Similarly with the Euler equation [69], you have acceleration of the first-order contribution, acceleration of the second-order contribution. Also, the product of the first-order deviation in density and the first-order deviation in velocity is a second-order term and I have to hang onto it, I cannot throw it away as was done in the linear problem.

Similarly here, this is an important term, $(\mathbf{v}_1 \bullet \nabla)\mathbf{v}_1$. That term will represent the convective part of the nonlinearity – the self-Doppler term. If you multiply that term by the ambient density, ρ_o, you generate a second-order term. Both the first-order contribution to the pressure gradient and the second-order contribution must also be included.

Similarly, one must expand the equation-of-state to second order. The first p_2 term in [70] is just the equilibrium sound speed squared times the second-order density variation, ρ_2. There is also a contribution from the second derivative of the pressure with respect to density times the square of the first-order density variation, ρ_1^2, if I am taking everything at adiabatic conditions.

Transparency #27 shows the relative strengths of the first- and second-order terms.

If we now return to Transparency #28 and do the same manipulation I employed to create the first-order wave equation, I produce a second-order wave equation [71]. Instead of it's being homogeneous; there is the Laplacian of a term that came from the convective part and a term that came from the second-order part of the equation-of-state. Lo and behold, we are back to convection and equation-of-state as the "drivers" of nonlinear wave distortion!

In fact, remember $\partial p/\partial \rho$ at constant entropy is sound speed, so $\partial^2 p/d\rho^2$ is just the density derivative of the sound speed, which is what you see in [72]. It is the same term I wrote way back in [10]. The term in parentheses in [72] is just Γ, the Grüneisen coefficient. What we find, if we do the hydrodynamics correctly to second-order, is a wave equation for the second-order terms (in this case I have chosen density but pressure would have been fine) that is driven by quadratic combinations of the first-order sound fields.

That new approach leads us to a different interpretation; another complementary picture of what is going on to produce nonlinear distortion. It is a very powerful and a very useful picture. It gives us another mechanism to explain what is going on and also provides a nice way to take a look at the effect of two waves of different frequencies and different wavelengths, when they interact in a nonlinear fashion. So far, we have only investigated self-interaction. That is why I am taking you down this third road.

[Transparency #29 – Parametric Array]

If we go back and reproduce what we did before, with just a single wave distorting, we can take a cosine-propagating wave of peak amplitude ρ', we can square it, and generate a constant term and then a "2f" term as shown in [73]. That is the term "driving" the nonlinear wave equation [72]. We can interpret quadratic term as an array of virtual sources. The virtual sources are created by the co-linear interaction of these waves squared; a whole bunch of little loudspeakers. We are generating sound at a frequency, ω, but now in one-dimension, along this column, there are a whole bunch of loudspeakers that have a source amplitude that is determined by that operator on the right-hand-side of [72]. Since $\omega/k = 2\omega/2k$, the sound from the virtual loudspeakers have the same phase velocity as the propagation velocity for both the fundamental wave and the second harmonic component, so they behave as an end-fire array for the second harmonic.

That is another way of looking at the initially linear growth we saw first in the Earnshaw solution [44]. You start with one loudspeaker. When its wave reaches the second loudspeaker, they add in phase, making the amplitude of the second-harmonic component grow with distance. The sum of the first and second loudspeaker outputs reach the third loudspeaker and add, in phase, to its output. This is also a good model for linear growth of the second-harmonic distortion component.

[Transparency #30 – Nonlinear (but Co-linear) Wave Mixing]

But why stop there? We have reproduced the Earnshaw solution [44]. What happens if we take two waves of different frequency and different wavelength? Let's just choose them to be not very different, so that the frequency of wave No. 2 is slightly less than the frequency of wave No. 1. The difference in their frequencies is much smaller than the frequency of either one.

What we envision here is a loudspeaker that is generating two frequencies, or possibly two different loudspeakers, at the same location, generating two different frequencies. Those frequencies are interacting within the medium. We have this quadratic term [72] that is "mixing" the two waves. The squared terms in [75] produce harmonic distortion of the "pump" waves. Then from the product of two terms with different frequencies, and trig identities, we have two additional terms that represent "virtual sources" at the sum frequency and the difference frequency.

The medium is generating distortion of the first pump, distortion of the second pump, and a wave that is at the sum of those frequencies and a wave that is at the difference of those frequencies. All of those waves have the same phase speed [76], so these are all driving that second-order wave equation in "geometrical resonance," producing a virtual end-fire array.

The difference-frequency term happens to be a particularly interesting. The reason it is particularly interesting is that the length of this virtual array is determined by the attenuation of the pump frequencies. Let's say the pumps are at 100 kHz and 102 kHz. The length of the array that is generating that difference wave, which is one of the four products that is being generated in [75], is determined by the attenuation of those high frequency pump waves at 101±1 kHz.

You can make a very long end-fire array for the low-frequency source and, therefore, a very directional low-frequency source from something that is really actually a very small high-frequency source. That has attraction for different applications. *(See article following Trans 30.)*

[Transparency #31 – Parametric Array Waveforms]

The picture, schematically, is this. You have two pump waves of fairly high frequency that are attenuating over some distance. Each wave is creating second harmonic components that are growing initially. Of course, since the pump waves are attenuating, and the second harmonic is at twice the frequency, the attenuation of the second harmonic is four times as fast [28]. The pump harmonic distortion is attenuating in a distance that is short compared to the attenuation distance of their parent (pump) wave. On the other hand, the difference frequency is growing linearly until this pump waves attenuate. What can you do with that?

[Transparency 31a – A New Kind of Sound Reproduction]

Here is an article from a recent German audiophile magazine that is really bizarre in the extreme. I think it is also technically a disaster. The technical discussion in that article is almost entirely incorrect, but I have got to show it to you, because it is so entertaining.

Imagine sitting in a concert hall. You and everyone else are enjoying some marvelous music and suddenly the artist addresses you. Only you can hear what the artist is saying. Incredible? No, mental hospitals all over the country are filled with people who hear other people talking only to them. (Laughter)

This works only as a joke.

Anyway, there are commercial units, and Bruce (Denardo) is going to demonstrate one of them tomorrow at NPS. It looks very much like the unit in this article. I do not know what the legal issues are. I know Andreas (Larraza) has one of the American units and this article describes a Sennheiser product, so I do not know if there is cross-licensing or patent infringement between one group and the other.

You all have now taken enough of this course already to know that the waveform in the figure is entirely wrong. The other pictures are kind of nice. They illustrate that end-fire array idea. Each "virtual loudspeaker" is delayed by a $\tau_1, \tau_2, \ldots \tau_n$. They are all adding in phase; that is kind of nice. Tomorrow, Prof. Denardo could choose one of you to receive a secret message: "Kill Garrett." (Laughter)

The best they could come up with in the article is to create the illusion of birds flying by. They suggest doing this by bouncing the directional sound beam off of a statue or off of a wall. Originally they were saying that this would be great for stereo, but you already know enough about parametric arrays to know that it is going to be lousy; the bandwidth is going to be terrible. The array length is fixed by the attenuation of the pump frequencies. If you have a very low frequency then you have a very short distance in terms of the low-frequency wavelength. If it is high frequency, then you have really a lot of sound being generated. It is just terrible in terms of fidelity.

They say, well, you can make talking statues and bounce it off of other things. Believe me, the Navy's use for this was much more intelligent. We will not talk about what it is.

[Transparency #32 – Phase Matching]

We talked about the interaction of co-linear waves in the absence of dispersion. I have said that the waves all have the same phase speed. I have assumed a dispersionless medium. If there is dispersion, then you do not necessarily get linear growth within the interaction region. After a certain distance, the "virtual sources" will start going out of phase. The nonlinearly generated wave will build up and it will shrink down and it will build up, it will shrink down; it will basically beat in space. That is the bad news, but the interaction can still be useful and interesting in a dispersive medium.

If there is dispersion, you are no longer restricted to co-linear interactions. The velocity of the 2f component is different from the velocity of the 1f component. You can use the "scissor effect" if the phase speed of the pump wave is less than the phase speed of the nonlinear product. You can take two beams and send them in directions that are not co-linear and, therefore, their phase velocity will be something higher than the phase velocity of the thermodynamic sound speed. As shown on the transparency, the $\mathbf{k}_1$'s will add up vectorially to something less than $2|\mathbf{k}_1|$, but the frequency will be 2ω and the phase speed will be the thermodynamic sound speed divided by the cosine of the half-angle. The transparency shows the full angle between them. The cosine is always less than one, so that speed is always greater than the thermodynamic sound speed, c_o [76].

If you have a dispersive medium where the sound speed increases with frequency, then you can take two waves that are not co-linear and still create a parametric array. In fact, that is part of what you need to do to understand the distribution of waves on the surface of the ocean.

There is very strong dispersion in deep-water gravity waves. That provides selection rules for the conversion of energy and the cascade of energy from low frequencies, the big rollers, down to the high frequencies in wind-drive waves on the ocean [see A. Larraza, S. L. Garrett and S. Putterman, "Dispersion relations for gravity waves in a deep fluid: Second sound in a stormy sea," Phys. Rev. **A41**(6), 3144-3155 (1990)].

[Transparency #33 - Dispersion]

If the velocity decreases with increasing frequency, then you cannot couple the pump waves to a nonlinear product. There is no way to phase-match, because [76] only leads to a higher velocity, not to a lower velocity. There is a forbidden region where the dispersion curve crosses over. Solid-state physicists like dispersion curves rather than sound speed as a function of frequency because they like to slap Planck's constant in front of everything and say that they are conserving momentum and energy but, of course, they are not. They are just doing phase matching.

One positive consequence of "downward dispersion" behavior is that if you find a medium where you can control the "bending" of the dispersion curve, as is the case for superfluid helium, then by changing the pressure, you can go from an allowed region to a forbidden region. In the forbidden region, you do not have to worry about dissipating waves as shocks, if you happen to be sending a very strong, highly focused beam of sound to make, say, an acoustic microscope. The people in Cal Quate's group at Stanford exploited this by going to the region where you get a decrease in sound speed with increasing frequency and suppressing the energy that they would lose due to shock formation.

Shock formation is not always necessarily a bad thing. These are two sonogram images [Note: these images could not included in the transparencies.] They are images of the same piece of tissue. This one clearly has less resolution contrast than the other one. The one with lower resolution was made by ensonofying the tissue with a wave at 2.5 MHz and looking at sound scattered by the tissue at the same frequency, 2.5 MHz.

Exciting the wave at 3.5 MHz and looking at 7 MHz made this image, the one with the higher resolution. This company was able to use the ordinary transducer in their imaging system at high amplitude creating second harmonic distortion to provide higher resolution by going to higher frequency and shorter wavelength. I will let the guys who are going to talk about medical ultrasound tell you whether or not it is a commercially important effect. It is definitely interesting.

[Transparency #34 – Mode Conversion in Solids]

There are two cases I would like to introduce where the scissors effect can be used to do something completely different. It is what we call mode conversion. In solids we know that there are two sound modes: There are longitudinal waves and there are shear waves. According to the laws of elasticity, the shear-wave speed, which is the square root of the shear modulus divided by the density, is always lower than the longitudinal wave speed, which is the square root of the Young's modulus divided by the density [77]. Shear waves will always be slower than longitudinal waves.

Since $c_S < c_L$, the phase speed of two non-co-linear intersecting shear waves can equal the phase velocity of the longitudinal wave. In aluminum, this occurs when the angle between the shear waves is 52°. In the mid-1960s, Rollins, Taylor, and Todd took a nice cylindrical piece of aluminum, put flat faces on one side that were separated by 52° and mounted shear wave transducers on the flats. At the opposite end of the cylinder, they received longitudinal waves. They were taking two slower waves and interacting them at an angle. The phase velocity, then, was resonant with the second-order longitudinal wave equation, and they were able to convert one mode into another mode by nonlinear effects, not by bouncing them off of boundaries, which is another way to convert energy between otherwise distinct modes.

[*Transparency #35 – Mode Conversion in Superfluid* ^{4}*He (HeII)*]

A similar type of mode conversion was demonstrated in superfluid helium, where, again, I mentioned there were two sound speeds. There was a slow thermal wave and there was a fast compressional wave. The slow thermal wave is known as second sound and the ratio of the speed of first sound to the speed of second sound, is a large number. The ratio is temperature dependent below T_λ, but typically it is about 10. Therefore, the angle of interaction for nonlinear mode conversion is almost anti-co-linear.

The measurement of nonlinear mode conversion from second sound to first sound was the topic of my thesis experiment at UCLA. The best way to get two plane waves to interact over a long distance at a precisely controlled angle is to generate the waves as higher-order (non-plane) modes in a waveguide of rectangular cross-section. By accurately controlling the ratio of the drive frequency to the cut-off frequency for the higher-order waveguide mode, you could finely control the interaction angle. The interaction angle is given as the square root of one minus the square of the cut-off frequency divided by the driving frequency [80].

[Transparency #36 – Resonant Mode Conversion in HeII]

Unfortunately, superfluid helium is expensive and difficult to store. You cannot easily get volumes that were very large. To maintain the superfluid, you have to do your experiments in a Dewar vessel (a fancy thermos bottle). My advisor, Professor Rudnick, remembered that his pal, Bob Leonard, had built a spiral waveguide as an anechoic termination for a probe-tube microphone [R. W. Leonard, "Probe-Tube Microphones, J. Acoust. Soc. Am. **36**(10), 1867-1871 (1964)]. What we did was we sent second sound waves bouncing up and down the spiral waveguide to interact over a meter inside a Dewar vessel that was only 6 inches in diameter. All of that useful interaction length just by winding up the waveguide as Leonard had done in air.

A pair of heaters, shown to the right of the spiral waveguide in the transparency, generated the thermal waves. If you look carefully, you can see that the heater elements are "shaped" like a sine wave to optimize the coupling to the first higher-order waveguide mode. The frequency of the drive was divided in half by a frequency divider that Prof. Keolian developed when he was a young child. (Robert and I were both at UCLA at the same time.) Since Joule produces the heating, I^2R, the frequency of the second sound is restored to that of the original drive frequency. You can take a look at the pressure (first sound) wave created by the nonlinear mode conversion process at the end of the waveguide opposite from the heater.

The graph at the bottom of the transparency [S. L. Garrett, S. Adams, S. Putterman and I. Rudnick, "Resonant Nonlinear Mode Conversion in He II," Phys. Rev. Lett. **41**(6), 413-416 (1978)] shows that the measurements (points) fall very close to the solid line that was calculated from the sound speed measurements of Heiserman, Hulin, Maynard and Rudnick [Phys. Rev. **B14**(9), 3862-3867 (1976)] that determined the thermodynamics of the superfluid [J. Maynard, "Determination of the thermodynamics of He II from sound-velocity data," Phys. Rev. **B14**(9), 3868-3891 (1976)]. The line has no adjustable parameters.

DR. SABATIER: What is on top of the heater?

DR. GARRETT: They are just two probably 0-80 screws where you could attach the wire. The heater is constantan wire that is wound back and forth in almost a sinusoidal profile, as well as I could do. I could do better "fine work" when I was 20, than I can now at age 50.

It is just two heaters. They are independent. The electrical current through the heaters is 90° out-of-phase, so the heating is 180° out-of-phase.

[*Transparency #37 – 2nd Review and Summary*]

So we have come to the end of the second portion of these lectures on nonlinear acoustics. We will finish up with non-zero time-averaged second-order effects, but I want to summarize first.

We decided to approach, in this second hour, the nonlinear effects by doing a perturbation expansion. The linearized hydrodynamics is correct only to the extent that the acoustic Mach Number is much, much smaller than one. When you can no longer neglect the Mach Number, then you have to include the second-order effects.

I started out by taking the sound speed and feeding it back to the first-order solution to the wave equation; that was the Earnshaw solution. We found that the second harmonic initially grows linearly with distance from the source. By doing a Fourier analysis of the geometrical construction of a distorting waveform, instead of the iterative picture employed to produce the Earnshaw Solution, we found exactly the same result as Earnshaw for the second harmonic growth, being linear with distance from the source, but we also picked up the third harmonic, the fourth harmonic, etc. We observed the required decrease in the amplitude of the fundamental (pump) wave, which had to provide energy for the generation of those other higher-order distortion products, so the Hargrove solution was a much more complete. But, again, the Earnshaw picture was nice, because you saw the second-harmonic generation through the effect of the first-order solution on the local sound speed (or at least I think it is nice).

Then we went down the "Royal Road." We said let's take the hydrodynamics and let's **not** linearize it. Let's keep all terms up to second-order. We found that the second-order contributions were generated by quadratic products of first-order terms. That approach led us to the end-fire parametric array model where you have all these "virtual sources" being generated by nonlinear mixing in the medium. If the medium was non-dispersive, all of the co-linear virtual sources added up in-phase, again giving us linear growth as we go away from the source. If there is dispersion, and it has increasing velocity with frequency, then you can spread the interaction angle and still meet the geometrical resonance criteria.

The nonlinearly generated "difference wave" was interesting because it allows us to produce a virtual low-frequency directional sound. In linear acoustics that would not be possible unless you have a "piston" with a diameter that was very large compared to the wavelength of the low frequency "difference wave." To get directionality, you need a source whose circumference is many, many wavelengths. At 100 kHz in air, you can have a source that is very directional but still not much bigger than your hand. The attenuation length of the 100 kHz sound will produce an end-fire line array antenna that is many, many wavelengths for the low-frequency difference tone.

Dispersion will lead to a de-phasing of that array and, in fact, you can use that concept to convert energy from a slow mode to a fast mode, or vice versa. Cherenkov radiation [L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, 1960), 357-359] is an optical effect that is similar in the sense that an electron traveling at a speed faster than light in some dielectric medium will emit electromagnetic waves as an angle determined by the ratio of the electron speed to the speed of light in the medium.

You now get another 10-minute break and then we are going to talk about acoustic levitation and the acoustic forces and torques produced by high-amplitude standing waves. This last topic should again be a useful introduction to some of the demonstrations Prof. Denardo will show you tomorrow at the Naval Postgraduate School.

[*Transparency #38 – Non-Zero Time-Averaged Effects*]

DR. GARRETT: I am going to shift gears and not look at the interaction of sound waves, either with themselves or with others. I want to introduce some fascinating non-zero time-averaged effects. As you know, if you look at the linear acoustic pressure, p_1, and you take its time-average [82], the integral of a sinusoidally varying function over a sufficiently long time will integrate to zero [83]. It produces as much positive pressure as it produces negative pressure.

Second-order term in the product [82] contain a constant that I have not been addressing in the earlier portions of this lecture. Of course, a constant has a time-averaged value that is non-zero; that is, $\langle p_2 \rangle_t$, the second-order part, when time-averaged, is non-zero. This can lead to substantial forces. There once was an acoustician named Hillary W. St. Clair. I do not really know too much about the guy, but he made some very impressive acoustic devices when he was at the US Bureau of Mines in Salt Lake City, UT [*e.g.,* "An Electromagnetic Sound Generator for Producing Intense High Frequency Sound," Rev. Sci. Inst. **12**, 250-256 (1940)]. In the *Review of Scientific Instruments,* back in 1941, he had a picture of a siren, a reflector, and three pennies levitated in space. Pennies are made from copper, they are eight times as dense as water, and they were being held up dancing in a sound field, so these forces can be substantial.

Those of you who have seen a Boeing 747 know that these forces can be substantial. There are unbelievable amounts of aluminum and upholstery that can get pulled off the ground because a fluid that is moving has a lower pressure than a fluid that is static. It is known as the Bernoulli effect.

You are probably familiar with that effect if you have a shower that has a shower curtain. When you turn on the shower you would think that the flowing water would push the curtain out, but, no, it always sucks it in. If you are a male graduate student, that curtain is disgusting. When that thing comes at you, it is terrible. (Laughter)

The point is that the water is driving the air inside the shower into motion and the air that is outside the curtain is not moving. The static external pressure is greater than the internal pressure, so anything that is as flexible as a shower curtain is going be moved by that small pressure differential. Everybody has experienced that. So even for fairly small velocities, if you have large enough areas and compliant structures, things will move.

[Transparency #39 – Non-Zero Time-Averaged Effects]

This effect of the Bernoulli pressure was discussed in Lord Rayleigh's *Theory of Sound* when he talked about the Kundt's tube. The Kundt tube is an air-filled is a tube that was used in the days before electroacoustics (much of this refers to the days before electroacoustics). The air resonances were excited when you would stroke a rod to create longitudinal vibrations at the end of the bar that would couple to the air in the resonator. You would adjust the length of the tube with a piston to make the bar and air resonances occur at the same frequency. Cork dust was placed in the resonator because it would be agitated by the high-amplitude standing wave. They did not have microphones and oscilloscopes in those days.

The cork dust, besides oscillating around, did a couple of other things that Rayleigh explains here. It would tend to form striations. You can see those striations in the lower photos. I was flabbergasted when I read the article that contained those photos, because I figured, man, nobody has used a Kundt tube for research purposes in a hundred years; it is a demo for lectures, but would anybody do research with a Kundt's tube?

Here it is, 1999, *The Journal of the Japanese Acoustical Society*. A guy was studying architectural bricks that are used for absorbing low-frequency sound and there was an apparent anomaly. An absorption peak was observed that was not at the Helmholtz frequency. The authors of the article went to great lengths to discover that it was the first standing-wave mode in the neck of the Helmholtz resonator. It is a concrete brick with a hole in it.

On the left, you can see the Helmholtz mode. You can tell that there is a large uniform gas velocity through the neck, since you see striations throughout the neck. When they went to the higher frequency, where the absorption anomaly occurred, low and behold, there were striations near the ends of the "neck", but nothing in the middle of the neck where there would be a velocity node for the first open-open, half-wavelength resonance. Striations at the ends but none near the center. This is still apparently a state-of-the-art research technique in Japan.

Lord Rayleigh in the *Theory of Sound*, Vol. II, Sec. 253b, described the striations. The reason you find these striations is that you have cork dust particles that are large enough that their inertia immobilizes them; so the gas flows by them. The cork dust does not move with the gas. This is not true, as Dick Stern points out, for smoke particles, which are so light that they move with the fluid flow. If the particles are massive enough that the flow moves around them, and if you have two particles aligned in the direction of the sound field, they will repel each other causing the striations. This is because there will be stagnation of the gas between the particles which block the gas flow. The pressure will be high in between them and everywhere else the pressure will be lower, because there is substantial gas velocity. The particles will repel each other, and that is what you see. The striations are lines of repelling particles. Of course, since the Bernoulli force depends upon v^2, it does not matter that the flow velocity is changing its direction twice per cycle.

On the other hand, if two particles are separated along a line that is perpendicular to the gas flow, you can see that the flow has to speed up to get between them. The Bernoulli Equation [84] guarantees that the pressure between them is lower and that lower pressure in-between will draw the particles together. That is also evident if you view the striations from the side instead of from above. The particles are stacked up together in the direction perpendicular to the gas flow. What you have are stacks of particles. They block the gas flow so Bernoulli pushes the stacks apart.

DR. ATCHLEY: What is the distance?

DR. GARRETT: Good question. I believe that basically it is the effect of the Laplace equation which governs the flow patterns in the absence of viscosity, $\nabla^2 v = 0$. There is only one scale length that can enter through the $\nabla^2 v$ term. I believe that the relevant scale is the height of the particle stack. The height of the stack is determined by the competition between the Bernoulli "stacking" force and gravity. How high they stack probably determines how far they separate. I think that is true, but it is intuitive, it is not algebraic. A striation spacing that is equal to the stack height is what you see if you look at the dust particles in the Kundt tube. The separation distance between striations is roughly the same as the height of the striations. That is what you would expect if the flow field is going to satisfy Laplace's equation.

[*Transparency #40 – The Rayleigh Disk*]

Instead of considering "point particles," let's say we examine an extended object in a sound field, like a disk. Bruce is going to demonstrate the forces and torques on a disk in a standing sound field tomorrow. The figure at the right is a resonator with a 1-inch loudspeaker at the right end and here is a circular disk suspended by a torsion fiber at the center of the resonator. The disk happens to be metallic, because the angle the disk makes with the axis of the resonator is measured by the five sets of electrical coils that surround the resonator tube [S. Garrett, "Butterfly-valve inductive orientation detector," Rev. Sci. Inst. **51**(4), 427-430 (1980)].

Let's say that the disk is at some angle with respect to the axis of the resonator and therefore to the oscillatory gas flow, as shown in the Figure 54a from *Theory of Sound,* if we were looking down toward the disk along the direction of the torsion fiber. When the gas is moving to the left, as shown by the arrows, there is a stagnation point on the upwind side between A and B. The gas has to speed up to get around the top at point B and travel quickly past point P to rejoin the flow. Similarly, when it leaves, the gas is coming around the other way from C and meets with the flow from P to create another stagnation point on the side of the disk opposite Q. At the stagnation points where there is no flow, there is high pressure. Where there is higher speed flow, there is low pressure. Therefore, the sound wave produces a moment on the disk that will produce a torque that tends to align the disk perpendicular to the fluid flow direction.

That torque is given by [85]. It involves, of course, the Bernoulli pressure so it is proportional to ρv^2. The torque depends on the angle, because when the disk is perpendicular to the flow, there is no torque on it; the stagnation is uniform across it.

When the disk is aligned with the flow, there is also no torque on it, so [85] has to be proportional to 2θ, where θ is the angle between the normal to the disk and the direction of the fluid flow in the absence of the disk. Of course, when the disk is aligned with the flow ($\theta = 90°$), the equilibrium of the disk is unstable. If it were to rotate just a little bit, it will slam over in a direction perpendicular to the flow ($\theta = 0°$). If it were to rotate just a little bit in the opposite direction, it would slam over the other way ($\theta = 180°$). If there were no restoring torque from the torsion fiber, the disk tends to align itself perpendicularly to the flow field.

That is why there are people in Hartford, Connecticut that were not killed by the windows that fell out of the Hartford Insurance skyscraper. That Hartford Insurance building had problems with the glass windows falling out. As we just saw, the windows did not come down like a knife and slice people into deli-sized pieces; the windows floated down like a leaves. Leaves do that because they are going to try to align themselves perpendicularly to the flow field.

Before the days of electroacoustics, the Rayleigh Disk was the most accurate way to measure the absolute amplitude of a sound field. You could measure the oscillatory particle velocity produced by the sound wave by measuring the torque on the disk. The torque rectified the high frequency oscillatory sound field since the torque depended on the square of the acoustic velocity. It did not matter whether the flow was going this way or that way.

[Transparency #41 – Bernoulli's Equation]

I do not want to spend too much time producing a hand-waving derivation of Bernoulli's equation. We can start from the hydrodynamic force (Euler) equation [86], including the nonlinear term (remember, this is obviously going to lead us to a v^2 term). The goal of the following manipulations will be to derive the Bernoulli equation by expressing the Euler equation as the gradient of some other quantity.

Instead of dealing directly with the $(1/\rho)(\partial p/\partial x)$ term in [86], I can include the pressure by use of the enthalpy function [87]. The enthalpy function is the internal energy of the gas, ε, plus pdV and Vdp. The internal energy is TdS-pdV, so if we substitute for the internal energy, it eliminates the pdV term in favor of TdS. If the sound wave is adiabatic, we lose the TdS term, since the entropy is constant in an adiabatic process. For a unit mass in a volume, V, the reciprocal of the specific volume is the density. So we can express the $(1/\rho)(\partial p/\partial x)$ using [86] as just the gradient in the enthalpy as shown in [88]. The $v\ (\partial v/\partial x)$ term is equal to one-half the derivative of v^2 by the product rule.

So all I have to do to get all of the terms in the form of gradients is to express the velocity as a velocity potential, ϕ, whose gradient I define as the velocity. Now I have all of the terms inside a gradient that is equal to zero [89]. The result [90] is the "strong form" of Bernoulli's equation: the time rate of change of the velocity potential plus $\frac{1}{2}v^2$, plus the enthalpy, is a constant. It is constant everywhere throughout the fluid, not just along a streamline. That is the big advantage of deriving the Bernoulli equation in the way I just did. Let me demonstrate the utility of that result in acoustics.

[Transparency #42 – Time-Averaged Pressure]

We take the strong form of the Bernoulli equation [90], and we expand it out to second order. This means the enthalpy, which is a thermodynamic function of the fluid, will also be expand to include both first and second order terms as shown in [91]. If the fluid entropy is held constant, $(\partial h/\partial p)_s$ is just $1/\rho$. The second derivative term is just derivative of $(1/\rho)$ with respect to pressure. It is one over the sound speed squared times $(-1/\rho_o^2)$, again by the product rule.

We end up now with an expression [92] that is correct all the way to second order in deviations from equilibrium. We can take the time average of the second-order pressure contribution, $\langle p_2 \rangle_t$, because all we are interested in are time-averaged effects since we are concerned about acoustic forces on objects that are so big that their inertia prohibits them from responding to pressure gradients at the acoustic frequencies. Of course, even "massive" objects can respond to steady-state gradients in the pressure.

We are using the fact that the objects we are interested in manipulating with sound fields have some mass to basically freeze out the AC part of their mechanical response, so we take the time average. When we time-average the derivative of the velocity potential, since it is sinusoidal, it vanishes. The equilibrium value of the enthalpy, h_o, becomes an irrelevant constant since we need pressure gradients to produce forces, so we can just amalgamate it with the other constant on the right-hand-side of [92]. We end up with two terms that have a non-zero time average that will contribute to the time-averaged second-order contribution to the pressure, $\langle p_2 \rangle_t$.

We ask ourselves, if we have a traveling wave, what is the time-averaged force produced by [93]? Since pressure in a traveling wave is the specific acoustic impedance, ρc, times the velocity, we can make that substitution for the quadratic velocity term in [93] and we find out that there is no pressure on a solid object immersed in an acoustic traveling wave field. Does that mean if we generate a traveling wave and we put an object in it, it will not feel a force?

It will feel a force, but I have just proved that in a traveling wave it will not feel a force. Does Garrett speak with forked tongue? Will there be a net DC force on it or not?

The answers to this apparent paradox is contained in the fact that you cannot put a solid object in a traveling-wave field and not have a standing-wave component produced by reflection of sound from the object. It is this sort of quantum mechanical paradox, where if you make the measurement, you create the force. If you put the object in there, there will be scattered component. The scattered component will combine with the traveling-wave component to produce a standing wave, so the object that scatters the traveling wave will feel a force. It is kind of an interesting result, if you choose to view it from that perspective.

[*Transparency #43 – Levitation Force*]

For a standing wave, you have a pressure that has cosine dependence in space, let's say, and a sinusoidal time dependence [95]. The velocity that corresponds to that pressure distribution is given to you by the Euler equation [96]. If we substitute those two results into equation [93], we end up with [97]. You now have $\cos^2$-$\sin^2$, which we know from trig identities that is cos (2kx).

We find that the non-zero time-averaged portion of the second-order force varies with twice the spatial periodicity, or half the wavelength of the first-order standing wave field. The second-order pressure is proportional to the first-order pressure squared or the first-order velocity squared. The pressure gradient will exert force on objects located within the standing-wave sound field. I am going to generate a standing wave and then put the object in; I am not going to cheat as I did when the object scattered sound and produced its own standing wave pressure field.

If there is a time-averaged pressure, the pressure difference across an object is the pressure gradient times the effective "thickness" of the object in the direction of the pressure gradient. A force on a disk is proportional to the product of a pressure times an area [98]. The product of the thickness of the object and the area of the object produces a term that is proportional to the volume of the object. For a sphere of radius, a, you end up a result [99] of exactly the same form: $4/3\pi a^2$ times a. That is, the volume of the sphere times the square of the pressure amplitude times the sin (2kx) divided by the wavelength, roughly.

Since the force on a sphere or a disk scales with the volume of the object and the mass scales with the volume of the object, assuming the object is made of a material with a constant density, then the levitation force on a solid object that is necessary to overcome gravity is determined, of course, by the wavelength (the shorter the wavelength, the more force you get). If you have an acoustic pressure amplitude squared that is greater than γ times the mean pressure times the density of the sphere, times the acceleration due to gravity, times the wavelength divided by π [100], you will levitate an object against the force of gravity. You will be able to hold objects in space using standing-wave sound fields.

MR. PORTER: Is this essentially radiation force?

DR. GARRETT: Yes, it is, that is one way to look at it, but, remember, the theme of this lecture is that understanding only one way is not enough.

[Transparency #44 – High Amplitude Standing Waves]

Levitation requires very large acoustic pressures. If you are levitating something with a density of, say, water, you are talking about 10% acoustic overpressures in standing waves, $p_1/p_o \cong 10\%$. Nothing I have presented thus far has prepared you to deal with nonlinear standing waves; everything I have told you so far has to do with traveling waves.

I now have to address the standing-wave issue, because I need these high-amplitude standing waves to do tricks with solid objects. One way to connect the traveling wave and standing wave problems has been suggested by Tempkin. He likes to think of a nonlinear standing wave in a resonator as being a traveling wave that keeps getting folded back on itself. Every time the wave hits the wall it is reflected, it comes back, it hits the wall, it is reflected, and that is how you build up resonance.

According to Tempkin, one can consider a resonator with some quality factor, Q, as being a very long traveling-wave tube that is QL long. The resonator is a traveling wave tube that is Q times longer than its physical length, L, for the purpose of calculating the nonlinear distortion.

If the resonator is what Prof. Atchley likes to call a consonant resonator, that is, a resonator with perfectly rigid ends, so that the harmonics are integer multiples of the fundamental (and that is true for a perfectly rigid resonator), then at high amplitudes, you have a shock wave slamming against the right wall, being reflected and slamming against the left wall. You just get this shock wave that keeps bouncing back and forth, it looks nothing like sinusoidal wave, just a shock front that bounces back and forth through the resonator.

In a consonant resonator, the second harmonic is also a resonance. The resonance of the cavity will enhance it. If the third harmonic is also a resonance in this cavity, it will get built up, and if the fundamental is a resonance of this cavity it will get built up, so you are building up the fundamental and you are building up all the harmonics.

DR. HARGROVE: It sounds as if you are using super position in a nonlinear context.

DR. GARRETT: Does it? (Laughter)

No, I am saying that the shock wave travels at the sound speed, which is true -- remember, the shock front travels at the speed of the zero crossing -- and it just propagates back and forth.

Professor Atchley has an article in *The Encyclopedia of Acoustics* that analyzes consonant and dissonant resonators. He defines a de-tuning parameter [101], this h_n that has nothing to do with the enthalpy. If you take the frequency of the $n^{\underline{th}}$ harmonic, subtract from it n times the frequency fundamental, and divide by n times the frequency of the fundamental, you obtain a relative frequency deviation parameter to quantify the de-tuning of the harmonics.

The effect of the de-tuning depends on the Q of the resonator. Since two modes are involved in the de-tuning, [102] creates an average Q for the fundamental and whatever harmonic is being examined. If the width of the resonance is within the overlap, then you have a consonant resonator and you have distortion problems. If not, you will be less susceptible to shock formation within the resonator.

MR. PORTER: My question goes back to traveling waves do not put a pressure or a force on an object. If I were to have an incident wave on a solution that has particles in it and there is noticeable streaming that is taking place, you are saying it is due to a standing wave --

DR. GARRETT: Ah, Tyrone just raised the stakes.

MR. PORTER: -- or is that due to attenuation in the fluid that leads to --

DR. GARRETT: There is another nonlinear effect that is on my list of topics I was not going to discuss, but I would be glad to talk about in the discussion section tonight. In fact, we have a really interesting perspective Roger Waxler has on streaming, because he has developed a unified view of streaming and radiation pressure and I have a partitioned view.

If a fluid has viscosity, there are dissipative effects that can absorb energy and momentum. The energy appears as heat but the rate of momentum loss must show up as a force on the fluid. That force creates fluid "streaming" that can drag along a particle immersed in the fluid. It is a very nice question, but we will save it for the discussion session this evening, because it involves a nonlinear effect that I have not prepared and do not have time to describe. It is a very important nonlinear effect.

[Transparency #46 – Nonlinear Standing Wave Shapes]

These are some waveforms that can be created in a consonant resonator. The middle waveform, labeled (d) in the diagram, is driven exactly at resonance ($\Omega = 1.0$). You can see clearly that you have a shock front. I suspect that the small amount of high frequency "ringing" is due to the natural oscillation of the pressure transducer diaphragm. The resonator has a circular cross-section and is as close to being consonant as they could make it. The top three waveforms: (a), (b), and (c) are driven below the resonant frequency ($\Omega < 1$). Waveforms (e), (f) and (g) correspond to excitation above the resonant frequency ($\Omega > 1$). You can see at resonance you have a really well-defined and very well-developed shock front.

DR. HAMILTON: Is that computational or an experiment?

DR. GARRETT: I think it is an experiment.

[Simultaneous discussion among participants.]

DR. GARRETT: These shock waves would limit your ability to levitate solid objects in sound fields and they would limit compression ratios for sonic compressors.

[Transparency #47 – Standing Wave Shock Suppression]

As we know from the parametric end-fire array picture, if there is dispersion, you can undo the distortion process by adding distortion components that are out-of-phase with distortion products created closer to the source. There are various ways of producing dispersion. You automatically get dispersion from thermoviscous effects, but you can intentionally put in dispersion by changing the resonator geometry. If you were to make the resonator a horn-shaped device, instead of a straight cylinder of uniform cross-sectional area, then you have a wave equation that explicitly involves the position-dependent cross-sectional area S(x) [105]. For an exponential change in cross-sectional area [103], when you plug the pressure wave [104] into this wave equation, you find that the phase speed, c_{ph}, is a function of frequency [106].

The phase speed depends on how close you are in frequency to the cut-off frequency, f_{co}, for the horn. The cut-off frequency is determined by the flare constant, h. If the flare constant is small, meaning the horn flares out very quickly, then you have a high cut-off frequency. Below cut-off you have no propagation along the horn. Above cut-off you have a phase velocity that goes from infinity at cut-off to the thermodynamic sound speed well above cut-off.

You can control the superposition of the harmonics geometrically. One way to look at shock wave suppression in standing wave resonators is to say that there is geometrical dispersion in the resonator. That approach is the one favored by a company called Macrosonix www.macrosonix.com, located in Richmond, VA. Macrosonix is a company that was founded to commercialize a nonlinear acoustic device that is a sonic compressor.

[*Transparency #48 – Waveform Shaping*]

Macrosonix makes their sonic compressor by shaking the entire resonator, shown here schematically as a trapezoid. The resonator cross-section is intentionally shaped so that you do not get shock waves but the "peaked" waveforms shown below the diagram. They shake the entire resonator to generate high-amplitude standing waves. At the small end of the resonator they have two sets of "flapper valves" with opposite flow directions. When the pressure within the resonator exceeds ambient pressure, the inflow valve is closed, and the outflow valve is forced open by the excess internal gas pressure. Compressed gas exits the resonator.

On the subsequent half-cycle, the pressure at the valve end is lower than ambient. The inflow valve is sucked in, allowing fresh gas to enter the resonator and be compressed and sent out again during the next half-cycle. This "acoustic ratchet" is a nice trick. Due to the international ban on production of CFCs legislated by the Montreal Protocols on Substances that Deplete Stratospheric Ozone; chlorofluorocarbons (CFCs) can no longer be used for refrigeration. This has caused problems for the refrigeration industry, because the substitute chemicals, the HFCs, which have no chlorine, are difficult to combine with the lubricants that are required to keep the refrigeration compressors from wearing out.

CFCs are compatible with hydrocarbon lubricants. That is why the compressor in your refrigerator will last 20 or 30 years. You can put in CFCs, and load up the compressor with just ordinary crankcase oil. The compressor will be bathed in oil and the oil will not degrade the CFCs. Everything will work perfectly for decades, except, of course, the Earth's stratospheric ozone layer, that has protected this planet from the harmful effects of ionizing ultraviolet radiation.

The HFCs are not compatible with hydrocarbon lubricants. You have to use very expensive esters to lubricate the pumps used with the new HFCs like R-134a. The people at Macrosonix wanted to build a refrigeration compressor that has only flapper valves, which are basically reed valves that do not require lubrication. You get away from the lubrication problem and you can still keep the refrigeration industry happy doing vapor-compression cycles.

Macrosonix took this idea of tailoring the waveform to new heights by using the geometrically induced dispersion to optimize the standing wave shape for compressor applications. You would like to have very high-pressure peaks but, of course, you can never produce negative pressures. This suggests that you would like to produce a waveform that has a fairly flat bottom but a fairly high-pressured peak. That shape allows you to push out gas at very high pressure and suck in the gas at very low pressure over a longer time interval during each cycle.

The desired waveform is shown on the transparency. Here I do have a computer simulation of the waveform. The dots are the measurements and the agreement between the two is excellent. This idea of shaping resonators to suppress shock formation has been brought to a very, very high level of sophistication by this company. The ratio of the peak pressure to the minimum pressure in this figure is three-and-a-half. I believe that have achieved pressure ratios close to ten in resonators without valves.

[*Transparency #49 – Modal Anharmonicity (Dissonance)*]

Three of their resonators are shown at the top of this transparency. Below is a spectral plot of the fundamental and the harmonics. You can see in the figure caption that Macrosonix has a particular way, which I think is quite a good way, of non-dimensionalizing the resonance frequencies. Using the total length of the device, they create a "reference frequency" by dividing the sound speed by twice the total length. Its fundamental frequency is 1.8 times this reference frequency. You can see the overtones are in no way harmonically related to 1.8: 2.2, 3.1, 4.2, and 5.1. By doing that, they can sum the different distortion components with the proper phase producing the desired waveform. You can do a Fourier analysis of the waveform on the previous transparency and ask what frequencies and what phases are required, and then calculate the resonator shape, just as Anthony did in his introductory lecture.

DR. WAXLER: How do you get a periodic waveform if you add up a bunch of different frequencies? I mean, if you are going to have some period of repetition in-between, it is not going to be periodic.

DR. GARRETT: I do not have a glib answer.

PARTICIPANT: It is driven periodically, so the response is going to be periodic.

DR. GARRETT: At the linear level, that is true.

DR. KEOLIAN: Even here, too. Harmonics are fundamental....

[Simultaneous discussion.]

DR. GARRETT: The distortion products occur at frequencies which are integer multiples of the fundamental that is designated Ω_1 = 1.789 times the reference frequency in the figure. They shake the resonator at a frequency of Ω_1. The higher modes (eigenfrequencies) of the resonator are shown as Ω_2 through Ω_5 on the spectrum. The higher-mode frequencies act to pull the nonlinear distortion products, which are generated at integer multiples of the drive, $n\Omega_1$, away from producing a simple shock front to producing the desired waveform which has a "fundamental" frequency of Ω_1. When you take the Fourier spectrum of the desired "pointy" waveform, its Fourier components are obviously at integer multiples of Ω_1.

[Transparency #50 – A Second Watch]

I would like to introduce another way of looking at the use of geometrical perturbations of the resonator shape to detune the overtones because I can combine the new approach with another concept, which will be the last topic of the lecture. We can vary the cross-sectional area of a uniform resonator by putting a rigid piece of material in the resonator at different locations. In the upper figure, I have placed a rigid cube of material that is small compared to the wavelength at the rigid end of the resonator of uniform cross-sectional area is S. The length of the resonator is L. The volume of the resonator, SL, is much larger than the volume, V, of this obstacle.

If I put that obstacle at the rigid end, the obstacle is going to raise the frequency of the fundamental half-wavelength mode of the resonator. The frequency of that fundamental mode is the sound speed of the gas inside the resonator divided by twice the length of the empty resonator, $f_1 = c/(2L)$. One way to see that the obstacle is going to raise the frequency of the fundamental, is to recognize that in the fundamental mode of oscillation, the gas at one end is being compressed while the gas at the other end is being expanded. The near the center, the gas is oscillating back and forth rather rapidly.

By excluding some of the volume at one end with an incompressible obstacle, the gas spring at the end with the obstacle is stiffer than the gas spring at the other end, because the obstacle excludes some of the gas volume. If the incompressible solid was made of wax, and I tilted this resonator into a vertical orientation, I could heat the end and melt the wax. The volume of the wax would not change, but it would flow to create a plug of uniform thickness at the end, as shown in the lower figure. The plug reduces the length of the resonator, which again has a uniform cross-section, by a length, V/S. The effective length of the resonator is shortened and the new resonance frequency is increased to $f_1^p = c/(2L_{eff}) > f_1$. Since all I am doing is excluding volume, the plug is going to have exactly the same effect on the frequency as the cubical obstacle as long as both were small compared to the wavelength of the sound.

You can see that the resonator length will be reduced by the thickness of this plug. By putting an obstacle at the pressure antinode, the fundamental mode frequency is shifted to a slightly higher frequency. Having the obstacle at a pressure antinode will raise the frequency of the mode.

[*Transparency #51 – Resonator Obstruction Model*]

If I take the same obstacle, before it was melted, and put it in the middle of the resonator, then the gas will have to move faster to produce the same pressure amplitudes as the ends of the resonator. The obstacle is now causing the flow streamlines to go around it. This increases the flow velocity and increases the kinetic energy. By Rayleigh's method, it means that the same obstacle is lowering the frequency of the fundamental mode. Rayleigh says that the frequency of an oscillator is proportional to the square root of the ratio of the potential energy to the kinetic energy. If the potential energy stays the same, and the kinetic energy increases, the frequency will have to decrease. The obstacle located at the center of a resonator will lower the resonance frequency of the fundamental and the same object at a rigid end will raise the frequency of the fundamental.

If we consider the second harmonic mode of the empty resonator, what we would have a pressure antinode at the center of the resonator, as well as having pressure antinodes at both ends. For the second mode, even though the obstacle is in the same position at the center, acoustically it is at the pressure antinode, so the second harmonic will be increased in frequency. Does everybody buy that? The second harmonic is just like having two separate resonators with a rigid wall at the center.

I cannot tell you as easily as I did for the case where the obstacle was at a pressure antinode what the downshift is for the fundamental mode when the obstacle is located at a velocity antinode. That is because the magnitude of the frequency shift depends on the shape of the obstacle as well as its volumes. What you can appreciate, in at least a qualitative way, is that if I have an obstacle at the center of the resonator of otherwise uniform cross-section, the obstacle will drive the fundamental frequency lower and drive the second harmonic frequency higher, so the presence of the obstacle insures that the second harmonic frequency is not exactly twice the fundamental frequency. The obstacle has made a consonant resonator into a dissonant resonator. That is another way to look at this geometrical dispersion.

[Transparency #52 – Two Double Helmholtz Resonators]

The obstacle in the middle of the uniform resonator starts to transform the resonator into a geometry that is known as a double Helmholtz resonator. If I make that obstacle bigger and give it streamlined contours, what do you see? You see the neck of a Helmholtz resonator that connects two volumes.

What you see in the upper figure is exactly the same double Helmholtz resonator shape, except there is a lot of thermoacoustic hardware thrown in. Besides the two loudspeakers at the ends of the volumes, there are two pairs of heat exchangers and two stacks in the two volumes. It is thin and empty in the U-shaped neck portion. One purpose of that resonator shape is to make the modes of the resonator dissonant instead of consonant. For thermoacoustic refrigeration, we like nice sinusoidal waveforms and certainly do not want to dissipate acoustic energy by generating shock waves.

Shown at the bottom is the limiting case of that double Helmholtz geometry. This is a very large double Helmholtz resonator that weighs about a ton and is almost seven feet long. Tom Gabrielson designed it. It has held as much as 900 psi of pressure. With half that amount of gas pressure, the stored energy in the gas is as much as the energy in 3.5 pounds of TNT. It is a monster. Again, it is a thermoacoustic device that wants to maintain sinusoidal waves, so we need to suppress the shocks. By putting in an obstruction or necking-down the center section, we de-tune the fundamental from the overtones.

[Transparency # 53 – Ehrenfest Adiabatic Principle (EAP)]

I said that if we have another way of looking at geometrical dispersion, we are going to get another insight into the levitation effect. To get that insight, what we have to do is apply another important theorem in mechanics known as the Ehrenfest Adiabatic Principle. It says that the energy contained in some mode, the n^{th} mode, divided by the frequency of that mode, f_n, will remain constant when you change variables in the system slowly [107]. Slowly is the "adiabatic" part of the principle. The energy-to-frequency ratio is only an invariant if you make the changes to the resonator on time scales that are long compared to the acoustic period.

One simple example comes from the blues guitar style. If you take a guitar string and do a Claptonesque kind of bend to increase the tension in the string, the frequency of the note will increase. You did work on the string by stretching it. You did that work slowly because the tension increased over about a half-second. If the note had a frequency at around one kilohertz, then the string vibrated about five hundred times while you were doing work against it, so it was definitely an adiabatic stretch. By doing work on the string, you raised its energy. But the Ehrenfest Adiabatic Principle required that the energy-to-frequency ratio was not changed. The energy increased, so the frequency had to increase also.

Unfortunately, this is a result from classical mechanics, as it was known in the $19^{\underline{th}}$ century. This form in [108] is actually recognized by a lot more people these days than the version in [107]. The version in [108] is known as Plank's law: the energy of a photon is a constant times its frequency, but it is just [107], where the constant has been given a name.

All of you have seen this, but you are saying, "What?" The energy over the frequency is an invariant? What?" But it is. If that is the case, then it can tell us something about levitation and stability and it can relate levitation forces to shifts in the resonance frequency of the resonator used to levitate objects using intense sound fields.

[*Transparency #54 – Work Against Radiation Pressure*]

Before I try to relate adiabatic invariance to acoustic levitation, I want to convince you that in the simplest case, adiabatic invariance makes sense. If I consider the resonator shown in the figure, driven in its fundamental mode, I have already told you that the pressure near the center is going to be lower because the velocities are high and the Bernoulli effect is going to reduce the mean pressure down. But it is a closed resonator. The mean pressure is lower in the center, from the equation-of-state; the density at the center is lower also. Since the resonator is closed, we still have the same amount of gas inside, so the density had better be higher at both ends. Again, the higher density at the ends corresponds to a higher pressure at the ends. The average density will remain constant, because no gas is getting in or out of the resonator.

If we take the frictionless piston shown at the left, and we move it inward a distance dx, then we have done work against the radiation pressure. If the resonator is oscillating in its fundamental half-wavelength mode, the time-average of the total energy is the sum of the kinetic plus the potential energy. Since those contributions to the energy are equal and out-of-phase, the time average of the total energy can be represented by just the maximum value of the kinetic energy, which is just the kinetic energy density, $\frac{1}{2}(\rho v^2)$, integrated over the entire resonator. The sin (kx) spatial variation in the velocity introduces another factor of 1/2. The total energy in the mode is the cross-sectional area, times $\rho v^2/4$, times the length of the resonator [109]. AL is the volume of the resonator that multiplies half the maximum Bernoulli pressure drop at the center of the resonator, where the velocity is the greatest.

From the earlier viewgraph [97], the time-averaged, second-order pressure, $\langle p_2 \rangle_t$, is proportional to the product of cos (2kx) within the resonator and $\frac{1}{4}(\rho v^2)$. When we push this piston in, we are doing work against that radiation pressure. We are pushing it slowly, so we are changing the resonator dimensions over a large number of cycles. The force on the piston, which is the pressure times the piston area, is multiplied by the displacement to give us the work that we have done with the piston against the radiation pressure [110].

Adiabatic invariance tells us that the change in energy divided the energy, $\delta E/E$, should be the negative of the change in frequency divided by the frequency, $\delta f/f$. That is exactly the result shown in [111]. The change in energy is the work we have put into the system by making a displacement of the piston against radiation pressure.

You can see in [111] that the only thing that is left after canceling like terms, is dx divided L. The A is cancelled, the ρ is cancelled, the v^2 is cancelled, and the 4's cancel, so we will end up with dx/L, which, in fact, is the relative change in the frequency, $\delta f/f_1$. If you shorten a resonator by dx, you will raise the frequency by δf.

It is pretty obvious, at least in this case, that the Ehrenfest Adiabatic Principle is obeyed. What does that mean for us?

[Transparency #55 – Levitation Revisited]

It means that we can revisit this levitation problem and look at it in a different way. I have shown you that by moving that block around in a resonator, the resonance frequency will depend on the position of that obstacle. For the fundamental mode, if the block is against an end, the frequency goes up. If the block is in the center, the frequency goes down. Frequency will be position-dependent.

If the frequency is position-dependent, the energy is position-dependent, because of the Ehrenfest Adiabatic Principle. Therefore, if the frequency is changing with position, the energy is changing with position. Forces can be expressed as gradients in energy. By simply measuring the resonant frequency as you move whatever object it is you want to levitate around in the resonator, you can figure out what the levitation force is without ever doing a levitation force experiment. That is a much easier experiment and it can be done at modest acoustic amplitudes.

You move it. You measure the resonance frequency. You move it again. You measure the resonance frequency. All you have to do is add the p_1^2 part and you have the levitation force.

[*Transparency #56 – Levitation Instability*]

So you have an object that is held in position by a standing sound field. The object is at equilibrium. What does equilibrium mean? It means the levitation force holding it up is equal to the gravitational force pulling it down. The equilibrium is stable so we can describe deviations from the equilibrium position by an harmonic-oscillator equation [112]. The object has mass and the sound field provides the stiffness. If the object moves up from its equilibrium position as shown in the figure on the right, then the levitation force will decrease and gravity is going to move it back down. If it moves down, it is going to get pushed back up because the pressure gradient is going to raise it up if you are in the fundamental mode.

But most acoustic levitation systems, all the ones I have seen, are operated at a fixed frequency that does not depend upon the position of the levitated object. When the object moves up and when this object moves down, it is changing the natural resonance frequency of the resonator. The motion of the object causes the resonance frequency of the resonator to get closer to, or farther from, the driving frequency. As the resonance frequency approaches the driving frequency, p_1^2 will get larger. If the change in the position of the object is pushing resonance frequency farther from the driving frequency, then p_1^2 will get smaller.

There is a change in the stiffness of [112], because the position of the obstacle changes the resonator tuning. If you are driving at fixed frequency there is going to be a change in the effective stiffness. No big deal, it is not a problem; it is just a change in the effective stiffness that depends upon the position of the object.

What is a big deal is that resonator does not respond to where the object is. Resonators respond to where the object was. As you know, a resonance takes time to build up to its steady-state amplitude. There is a characteristic time for a resonance, say, τ_n for a mode that has a quality factor of Q_n, and some resonance frequency f_n. The time τ_n, is the time that it takes for the sound field to catch up with the position of the obstacle.

So you not only have de-tuning but you have this de-phasing. The sound field at any given time is the field that would be near its steady-state amplitude when the obstacle was at a position at some earlier time τ_n. Let's take a look at the consequence of this retarded response. Let us assume that the tuning of this resonator is sharp; that is, if the frequency of the driver is higher than the resonance frequency of the resonator. When the object moves down, the resonance frequency goes up and moves closer to the driving frequency, so the pressure amplitude increases and the levitation force on the object also increases.

If the force increases when the object is at its lowest point, just after it reaches its lowest point, the levitated object starts moving up. It is moving up when the force is getting bigger because of the retardation. The peak in the force will come a little bit later than the time when the object is at the bottom of its motion and has no velocity. The force and the velocity will be in phase as it starts to move up, so the sound field will do a little additional work on this object. It will add to the energy of the small oscillation because of this retardation and the change in tuning.

When the object gets to its top position, if the driver is tuned sharp, then you are farther out of tune at the top of the oscillation, so the force is less than it would have been had there been no change in the tuning. The force of gravity will be slightly too large. Gravity is pushing down. When the object is at its highest point in the oscillation, it is about to go down, but the force is going to take a while to decay from that de-tuning and you will get that additional force of gravity applied to the object when it is moving down. Again, the force will be in-phase with velocity. That gives you a resistance term that is negative and the oscillations will grow in amplitude.

The point here is that if your driver is tuned above resonance frequency of the resonator, any oscillation in the position of the object around its equilibrium position will grow. Initially that growth will be exponential in time. This will destabilize the system. In fact, there is a nice report about how acoustic levitation control of objects on the Space Shuttle was a complete disaster. They would levitate the objects and the objects would go slamming up against the wall of the resonator. Since the goal of using acoustics to hold the objects away from the walls to do "containerless materials processing," slamming samples into resonator walls was *verboten*. The whole objective was to keep it away from the wall. That was the point at which I was called into the project as a consultant to the Jet Propulsion Laboratory in Pasadena, CA.

[Transparency #57 – Levitation Superstability]

On the other hand, if you tune flat, with the driver frequency lower than the resonance frequency of the resonator, the de-tuning/de-phasing process produces additional damping [US Pat. No. 4,773,266 (Sept 27, 1988)]. If you tune the system flat, exactly the opposite thing will happen – the system stability will increase.

If you tune flat, when the object moves down from its equilibrium position, the sound field is going to provide less levitation force. When the object starts moving up from its lowest position, the force and velocity will be out-of-phase and the resistive term, R_m, in [112] is positive, so you increase the damping in the system. In addition to the Stokes drag on the object from its motion through a viscous liquid, you have an active system that is going to oppose any motion whatsoever, but only if you tune flat rather than tune sharp.

What the people at JPL realized, when I explained how this all worked, was that they were always "tuning sharp" because they were tuning their levitation chamber exactly to the resonance frequency **before** they put their object in the resonator. When you put the object in, if it is close to the center where they levitate it, then you are tuned sharp, because the resonator just went flat.

At that point, we all ran down into the basement where the levitation laboratory was located at that time and you will see the movie of what came out of that new understanding. But, yes, STS-41B was a disaster in 1984, stuff was just slamming up against the walls, and nothing worked.

Let me show you that effect. I will show you the JPL levitation device, which is kind of a pretty one. [Video]

Here is your levitator -- it is actually a three-axis levitator. There is the Styrofoam sphere that is going to be levitated. You can excite standing waves in the y- direction, in the x-direction, and in the z-direction. We are going to be concerned with only the z-direction since the sphere will be lifted against the force of gravity. There is a little piece of window screen that keeps the sphere from falling all the way down to the bottom of the levitator.

At first you are going to see this thing tuned at resonance condition, which I already pointed out is really as little bit sharp. See the ball come up. We will crank up the power. It takes about 15 cycles for the oscillations of the sphere to almost settle down. See how it keeps dancing around?

Now we are going to use acoustic damping and viscous damping. You saw how long it took to settle out with only viscous damping. Here it goes. Two-and-a-half wiggles and it is stable. Notice it is a little bit lower in the resonator than it was before. This is because now they are not at resonance exactly so they are not going to generate as much pressure.

Now they are going to tune it intentionally sharp. You can count the oscillations all you want but that sphere is never going to stop oscillating. The sound wave is just going to keep pumping energy into those oscillations because of the de-tuning. It does not take too much imagination to realize that the effect will be even more destabilizing if there were no gravity.

[Transparency #58 – Final Summary and Conclusions]

All right, we have reached the final summary slide. Let's see what have I done to take you away from the volleyball court? We have examined the life cycle of a plane progressive wave. I have emphasized again and again that the linear approximation is not correct, because the fluid convection caused by the wave motion, combined with the change in the equilibrium sound speed produced by the wave modulates the local sound speed. We have expressed the local sound speed as the thermodynamic sound speed plus some coefficient, Γ, which only Seth Putterman and I call the Grüneisen coefficient, times the fluid velocity produced by the wave.

The sinusoidal waveforms of sufficient amplitude are unstable. They start out as a pure sine wave; they will accumulate harmonic distortion with increasing distance from their source. Eventually, for large enough distances and strong enough waves they will evolve into a stable sawtooth waveform. For distances beyond shock-inception distance, D_s, all you ever have is a sawtooth for sufficient amplitudes in one-dimensional propagation.

Once the nonlinear distortion processes produce the sawtooth, the linear dissipation mechanisms that you have grown to know and love in linear acoustics does not describe the attenuation of the sawtooth; something else happens. The gradient produced by the shock or the steepness of that wave front, controls the amount of energy that needs to be removed and by putting in more energy, all you do is make the wave front steeper. The wave has to produce bigger gradients to consume that additional energy, so there is the natural saturation process that limits the amplitude of sound waves.

I then tried to do a somewhat more formal treatment of the nonlinear distortion by using second-order perturbation techniques to solve the hydrodynamic equations. I think I skipped that viewgraph, but on Transparency #27 in your handouts, you will see that as the Mach Number becomes significant, you have to include additional terms in the hydrodynamics that you previously neglected in the linear case.

Using the second-order perturbation series, we found that the wave equation for the second-order contributions to the sound field is now inhomogeneous and the source terms for that inhomogeneous wave equation are proportional to quadratic combinations of the linear sound field variables. That result was used to explain the generation of harmonic distortion through a linear end-fire array source model. That model led to linear growth of the second harmonic distortion when there was phase-matching. If there is dispersion, or the waves are not co-linear, then other things can happen, like mode conversion.

[*Transparency #59 – Final Summary and Conclusions (Con't.)*]

I then talked to you during this last hour of the lecture about non-zero time-averaged nonlinear effects. Most of what I said during the first part was that we could use the Bernoulli equation that expresses pressure reduction due to the square of the flow velocity. The velocity squared is a positive definite quantity, so its time average will be non-zero. Those second-order time-averaged pressures will cause point particles to agglomerate and to separate. It will create torques on extended objects such as a Rayleigh disk and can be used to levitate solid objects against the force of gravity.

At that point we were dealing with standing waves instead of traveling waves. We saw that we could suppress that shock-wave formation by changing the resonator geometry, and the company that is trying to build sonic compressors had taken that geometrically-induced dispersion to the status of a real art form. Their horns and cones and bulbs were designed to tune up the phases of the different distortion components and produce custom wave shapes that not only are not shocked, but also may be tailored to a particular application.

I also introduced an interpretation of this geometrical dispersion in terms of little blocks that were occluding the channel at the center of resonators or excluding volume at the resonator ends. That model was combined with Ehrenfest's Adiabatic Invariance Principle to relate changes in frequency to the spatial gradient in energy that produces levitation forces. Viewed in the opposite way, the adiabatic invariance showed that the de-tuning, when coupled with the de-phasing, could increase or decrease the stability of that levitation point.

We have reached the end of this second lecture: Nonlinear acoustics made simple. These are my final words:

1. Know your limitations. When the Mach Number starts to become something bigger than one part in 10^4, 100 ppm, the results of linear acoustics may not apply.

2. If the Goldberg number is large, that is, the shock formation distance, D_s, is shorter than the exponential attenuation length, $1/\alpha$, you are going to have to consider nonlinear effects.

I restricted the mathematics, to make analysis of nonlinear phenomena as simple as possible, to one-dimensional systems. If you read the literature, if you look at what Mark Hamilton and others in the field publish, it is a lot harder. I cannot do it. The basic principles here are the same; the execution just gets to be extremely challenging. The beams can be divergent in three dimensions, and spherical spreading has to be taken into account. It takes really ugly mathematics to handle those problems, but there is nothing in there you have not seen in terms of the physics.

Hopefully, this extra effort in understanding these limitations of linear acoustics has been rewarded with some interesting new effects. Enjoy your volleyball game and thank you for your attention.

ACOUSTICS DEMONSTRATIONS

Bruce C. Denardo
Naval Postgraduate School

A transcript was not made of the Acoustics Demonstrations. Included in this document is a description of each demonstration that Dr. Denardo provided.

Dr. Denardo wishes to acknowledge the following individuals for their contributions to the demonstrations:

Robert Keolian – Pennsylvania State University
Andrés Larraza – Naval Postgraduate School

David Grooms, George Jaksha, Robert Sanders – NPS Staff

Contents of the follow demonstration descriptions:

1. Few-Degree-of-Freedom Oscillations
2. Mechanical Waves (Not Including Sound in Air)
3. Sound in Air

Oscillations

1. Uniform circular and simple harmonic motions

2. Spherical pendulum

3. Foucault pendulum model with round, I-beam, and square rods

4. Physical pendulums: two rectangular plates with holes

5. Torsional oscillator and mock demo of determination of G

6. Coupled oscillations: Lissajous pendulum, V-coupled pendulums, Wilberforce oscillator, and double pendulum

7. Relaxation oscillations of an RC circuit with a neon bulb

8. Electromagnetically coupled oscillators

9. Resonance pendulums

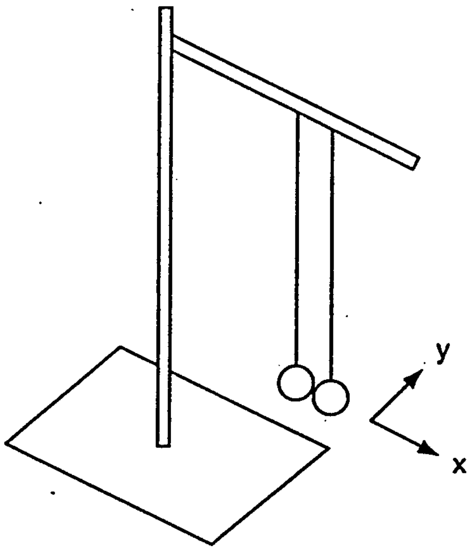
10. Keolian's driven oscillator: bent tuning curves and hysteresis

11. Parametrically driven pendulum: hand and loudspeaker

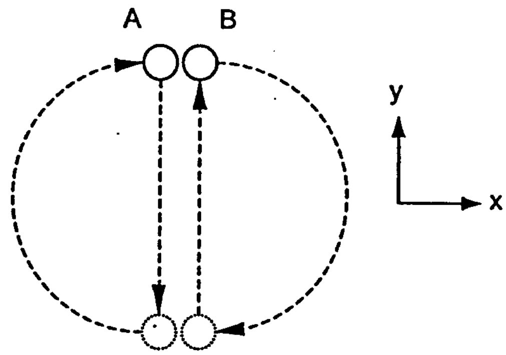
12. Parametric instability: spring pendulum and V-coupled pendulums

13. Phase locking: metronomes and organ pipe

OS.1 Uniform circular and simple harmonic motions



Two identical billiard-ball pendulums are side-by-side and touch in equilibrium.



Top view of the motion. The balls are pulled aside, and ball A is given an initial velocity in the x direction while ball B is initially held at rest.

The nearly elastic collisions of billiard balls are utilized in this demonstration. For small amplitudes compared to the pendulum length, the two balls are displaced equally in the y direction (see diagram above). One ball is held at rest while the other is given an initial velocity toward the first, which is released at the moment of impact. It is important that the collision be head-on. The incident ball is given a magnitude of the initial velocity such that motion is roughly circular. The motion of the system is then observed to repeat, as shown in the second diagram above. This shows that conical and planar pendulums of the same length have the same period for small amplitudes, which occurs because simple harmonic motion in a plane is the projection of uniform circular motion onto the plane. Hence, the position in simple harmonic motion varies sinusoidally.

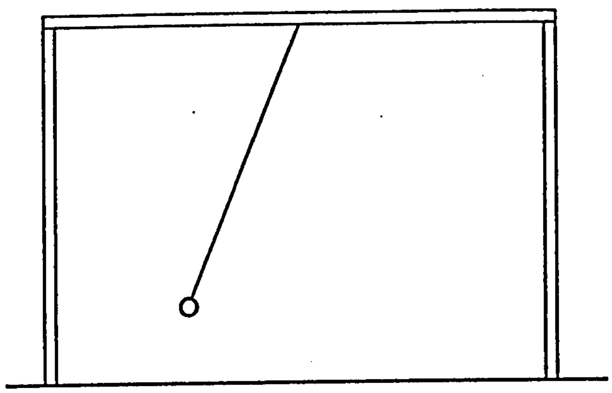
The magnitude of the initial velocity is not critical as long as the resultant amplitude is small. The motion is in general an ellipse, and the period is independent of the eccentricity, so the motion of the system repeats in this general case.

The colliding ball is eventually observed to have more and more forward momentum after each collision, which occurs because the collisions are not perfectly elastic. The amount of forward momentum slowly accumulates, and is apparent after roughly 6 collisions. After many collisions, the balls move together.

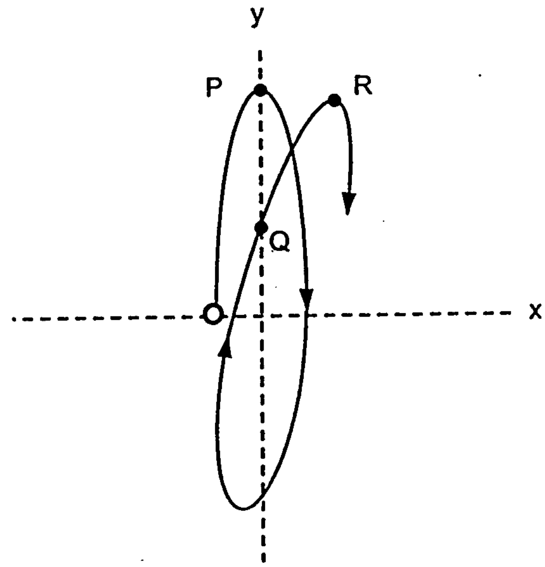
When the demonstration is done at higher amplitudes, the motion typically becomes unstable after several cycles. One reason for this is that the period of a conical pendulum *increases* with amplitude, whereas the period of a planar pendulum *decreases* with amplitude. Hence, the collisions cease to be head-on.

OS.2 Spherical pendulum

A spherical pendulum can be used to show how symmetry breaking leads to precession. For small amplitudes, Hooke's law holds and the general motion is seen to be elliptical. However, when one amplitude is large and the other small, the pendulum precesses in the direction of motion (*forward* precession).



Spherical pendulum, which oscillates both in and out of the plane of the figure. The trajectory lies on a spherical surface concentric with the support.



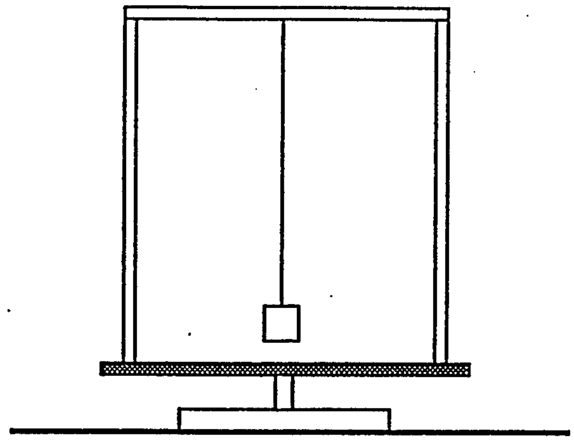
Top view of the trajectory of a spherical pendulum whose amplitude is initially small in the x direction and large in the y direction. Forward precession occurs.

The precession occurs because the period of a planar pendulum decreases for greater amplitudes (the oscillations *soften*). To understand how this causes the precession, consider the second diagram above. Because the motion is approximately linear, we can view it as a superposition of motions in the x and y directions. The x motion has a greater frequency, so this motion has gone though one cycle from P to Q, whereas the y motion requires a greater time, going through a cycle from P to R. Hence, forward precession occurs. For oscillations that harden, backward precession occurs.

Only two central forces yield orbits that close on themselves: inverse-square (Newtonian gravitation and Coulomb electric force) and Hooke's law (elastic deformations). Both orbits are in general ellipses. There is a symmetry or degeneracy here, because the orbital frequency equals the radial frequency. General relativity breaks the symmetry with effects that are not inverse-square. Mercury's elliptical orbit forward-precesses at a rate of 1.3 degrees per century. Of this, 0.012 degrees per century is due to a general relativistic effect. This additional precession was empirically discovered in 1845 by Urbain LeVerrier, who proposed that a new planet orbiting nearer to the sun was responsible. He named the planet Vulcan. It should be noted that more people today know of a planet by that name (from the Star Trek television series and movies) than of Einstein's achievement that accounted for the additional precession of Mercury.

OS.3 Foucault pendulum model

Vibrating rod pendulum clamped to a support which is attached to a rotatable platform. The plane of oscillation is observed to remain fixed as the platform is rotated.



In 1851, Foucault employed a pendulum to show that the earth rotates about its axis. Years earlier, he had noticed that the plane of vibration of a vibrating rod with one end clamped in a lathe remained fixed as the lathe turned. A universal ball joint with negligible friction, for which it is obvious that the plane of oscillation is fixed, is *not* required.

A Foucault pendulum at the north or south pole can be modeled by the system shown in the diagram. The restoring force for this oscillator is both gravity and the flexural stiffness of the rod, although what supplies the restoring force is irrelevant here as long as it is isotropic. The mass is set into oscillation in a plane, and it is observed that the plane remains fixed as the platform is slowly rotated. For an observer on the platform, the plane of oscillation appears to precess 360° for every revolution of the platform. Foucault performed his pendulum experiment not at a pole, but in Paris, France. What happens in this case? Such a situation can be simulated in the above model by arranging the rod to be inclined in equilibrium. The plane of oscillation is now observed to have precessed after one rotation of the platform. Examination shows that, for an observer on the platform, the plane of oscillation appears to have precessed less than 360° for every revolution of the platform. It can be shown that the angle of precession for one cycle equals the solid angle subtended by the axis of the pendulum (Hannay's angle).

The orientation of the pendulum is now returned to the vertical, and the demonstration is repeated. That the plane of oscillation remains fixed depends upon the isotropy of the oscillator. What happens for an *anisotropic* oscillator, which has two distinct frequencies in perpendicular directions? This is readily demonstrated with a pendulum having a plastic I-beam rod (available in hobby stores). An I-beam is stiffer for oscillations in the plane of the I rather than perpendicular to the plane. One of the modes of the pendulum is excited, and the platform is rotated slowly. The plane of oscillation is now observed to rotate with the table. This is an example of *adiabatic invariance*: if a mode of a conservative system is excited and one or more parameters of the system are slowly altered, the other modes remain unexcited. If the platform is not rotated slowly, the other mode is observed to become excited, and a Lissajous pattern results.

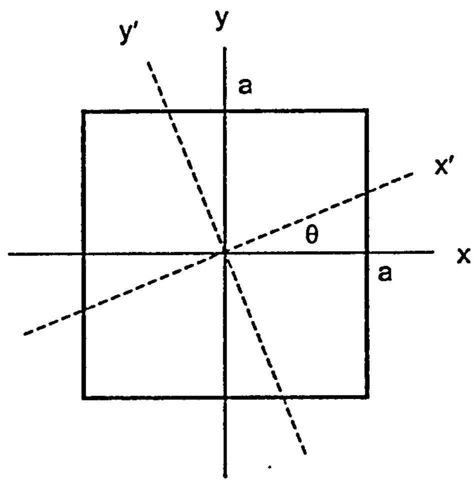
A rod of rectangular cross section would yield the same result as the I-beam rod. What happens for a rod with *square* cross section? This is not obvious. On one hand, it is natural to think that a square rod has a different stiffness when flexed perpendicular to a face compared to a diagonal plane (rotated 45° from a face), and so there should be two distinct frequencies of oscillation. On the other hand, oscillations of normal modes with different frequencies should be perpendicular to each other. The demonstration is done with a square rod, and the plane of oscillation is observed to remain fixed. This indicates that the flexural stiffness of the rod is isotropic. If the rod is flexed by hand, a difference in stiffness perpendicular to a face and to a diagonal plane is not detected. It can be shown (see below) that the flexural stiffness of a square rod is indeed isotropic.

The moment of force that results when a rod is bent is proportional to the integral of the square of the distance of a cross-sectional element from the midplane:[1]

$$\kappa^2 = \frac{1}{S}\int y^2\,dxdy\ ,$$

where S is the cross-sectional area and y is in the direction in which the rod is bent. Performing the integral for axes that are parallel to the faces of the rod (see diagram below) yields $\kappa = a/\sqrt{3}$, where 2a is the side length.

Geometry of the cross section of a square rod, and transformation of coordinates.

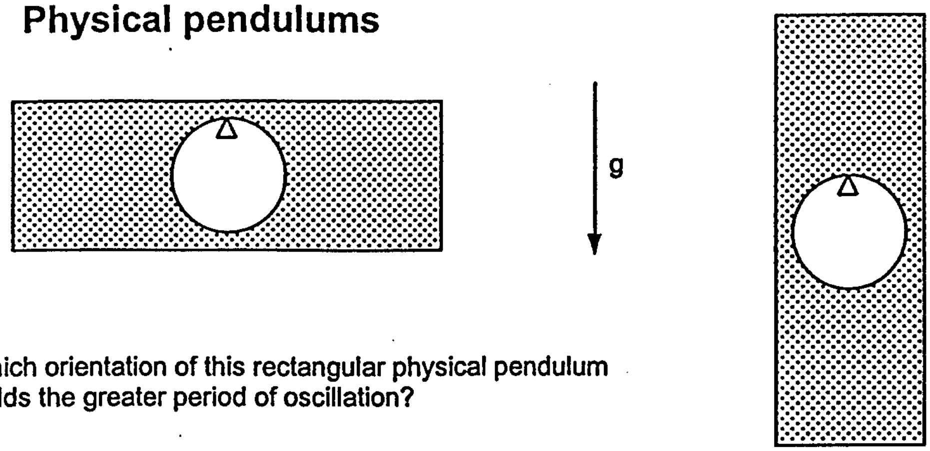


$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

We determine the value of κ when the rod is bent in the y′ direction (see diagram above) by transforming to xy coordinates. This is convenient because the limits of integration are simple in these coordinates ($x = \pm a$ and $y = \pm a$). Substituting $y' = -\sin(\theta)x + \cos(\theta)y$ and $dx'dy' = dxdy$ in the expression for κ^2, squaring the expression for y′, integrating, and simplifying, yields a value independent of θ: $\kappa = a/\sqrt{3}$ (the same value found above). This establishes the surprising result that a square rod has the same bending stiffness in any direction.

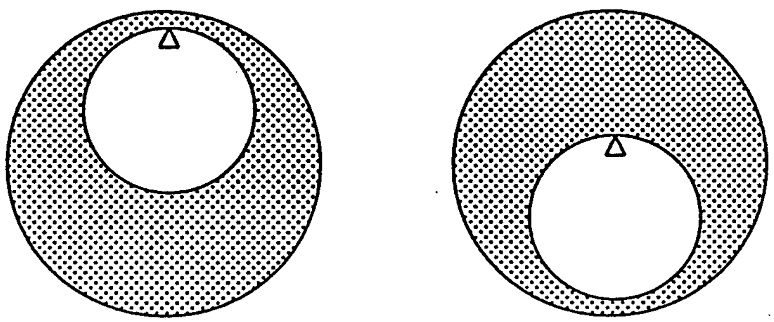
[1] Philip M. Morse, *Vibration and Sound* (Acoustical Society of America, 1981), pp. 151-153.

OS.4 Physical pendulums



Which orientation of this rectangular physical pendulum yields the greater period of oscillation?

Which orientation of this circular physical pendulum yields the greater period of oscillation?



These physical pendulum apparatus are machined from 1/4-inch thick aluminum. A brass knife edge clamped to an upright is used as a pivot. The pendulums have a number of interesting features.[1,2]

To determine the period, one must calculate the distance from the pivot to the center of mass, as well as the moment of inertia about the pivot. The center of mass is known by symmetry in the first case. How can the other center of mass and the moments of inertia be accurately calculated? The principle of superposition can be employed here. The first pendulum can be considered as the superposition of a rectangular plate of uniform density ρ and a disk of uniform density $-\rho$, and similarly for the second pendulum. The moment (mass times distance to center of mass) and moment of inertia of each pendulum are thus the *differences* of the corresponding quantities of the simpler objects.

How does the period of the rectangular pendulum change if the equilibrium orientation of the pendulum is rotated by 90^o? The distance to the center of mass, which is the radius of the hole, does not change. Surprisingly, the moment of inertia also does not change as a result of the parallel axis theorem, which relates the moment of inertia I about an axis is related to the moment of inertia I_o about the parallel axis through the center of mass: $I = I_o + Md^2$, where M is the mass and d is the distance from the axis to the center of mass. The period is the same in the two cases because d is the radius of the hole. (In fact, the period is independent of the location of the pivot along the circumference of the hole.) This equality is not obvious because the equilibrium orientations of the pendulum are geometrically different. They are, however, dynamically equivalent. The equality of the period for the two orientations can be verified by timing a large number (say, 10) of oscillations with a stop clock. Alternatively, a direct comparison can be made with two pendulums of identical shape but different orientation. Note that these demonstrations can be considered as experimental confirmation of the parallel axis theorem.

Another interesting demonstration can be performed with the two rectangular pendulums in the above diagram. For simplicity, both are arranged to have the same orientation. One is then rotated a small amount (roughly 5 to 10°) so that it is no longer perpendicular to the support. The pendulums are then set into oscillation with the same amplitude and phase. After some time, it is apparent that the phase of the rotated pendulum is gaining relative to the other pendulum. Hence, the rotated pendulum has a slightly smaller period. This occurs because the moment of inertia is slightly less.

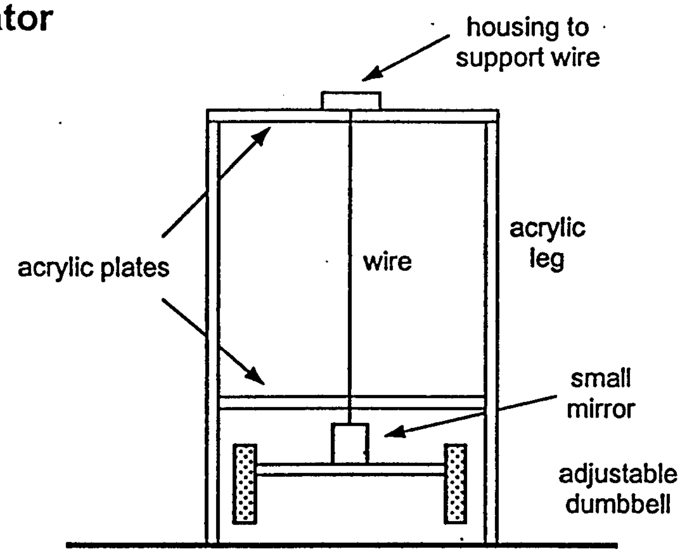
The circular pendulum can be inverted, and is observed to oscillate about this equilibrium orientation. From this stability, we conclude that the center of mass must lie inside the hole. By inverting the pendulum, both the distance from the axis to the center of mass and the moment of inertia about the axis decrease. However, the relative decrease of the first is much greater than that of the second, so the period increases. Indeed, as the distance to the center of mass of any physical pendulum approaches zero, the moment of inertia approaches a nonzero constant, so the period approaches infinity.

[1] Bruce Denardo and Richard Masada, "A not-so-obvious pendulum experiment," *Physics Teacher*, vol. 28, pp. 51-52 (1990).

[2] Bruce Denardo, "Demonstration of the parallel-axis theorem," *Physics Teacher*, vol. 36, pp. 56-57 (1998).

OS.5 Torsional oscillator

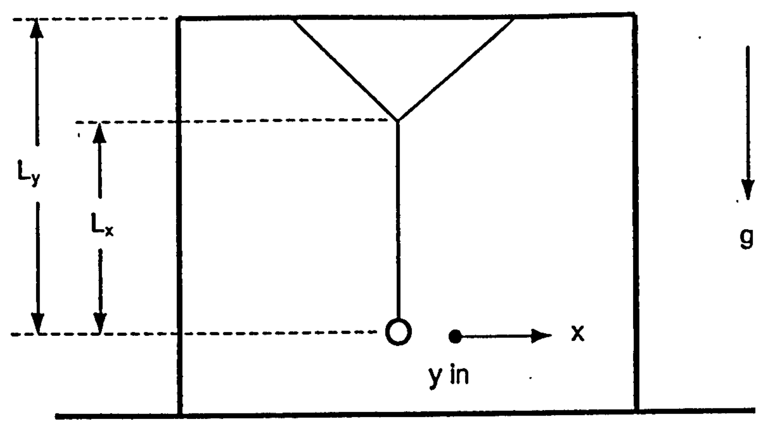
Torsional oscillator. After being initially displaced, the dumbbell rotates back and forth due to the shearing stiffness of the wire.



A dumbbell made of movable brass disks and an aluminum rod is suspended by a wire. The support structure is made of clear acrylic. After the dumbbell is angularly displaced from equilibrium or given an angular impulse, it undergoes oscillatory rotations. The disks can be moved nearer to the center to reduce the moment of inertia and thus increase the frequency. It should be noted that the mass remains the same in this case; it is the *distribution* of the mass that has been altered. In the absence of the small mirror (refer to the diagram), the disks can be brought together so that they touch the wire. The moment of inertia in this case decreases by roughly an order of magnitude, causing the frequency to roughly triple, because the period of a torsional oscillator is $T = 2\pi(\kappa/I)^{1/2}$, where κ is the torsional constant and I is the moment of inertia.

This apparatus can be used to illustrate Cavendish's experimental determination of the universal constant of gravitation G. A small mirror is glued to the horizontal arm, as shown in the above diagram. Laser light is reflected off the mirror to a distant wall. Very small motion of the torsional oscillator can be detected with this "optical lever." Cavendish measured the steady deflection of the torsional masses due to the presence of two large masses. Currently, the most accurate measurements of G are made by measuring the increase in torsional frequency when two large masses are brought near the torsional masses.

OS.6a Lissajous pendulum



Lissajous pendulum, where the ratio of the frequencies in the two directions is $f_x/f_y = (L_y/L_x)^{1/2}$.

As shown in the diagram, a ball (for example, a billiard ball) is suspended with string from a support. The frequency of motion in the x direction is $f_x = (1/2\pi)(g/L_x)^{1/2}$ and in the y direction is $f_y = (1/2\pi)(g/L_y)^{1/2}$, so the ratio of the frequencies is $f_x/f_y = (L_y/L_x)^{1/2}$. Possible dimensions are $L_y = 60$ cm and $L_x = 40$ cm, so that the vertex is 20 cm from the bottom of the support. The ratio of the frequencies is then $f_x/f_y = (3/2)^{1/2}$.

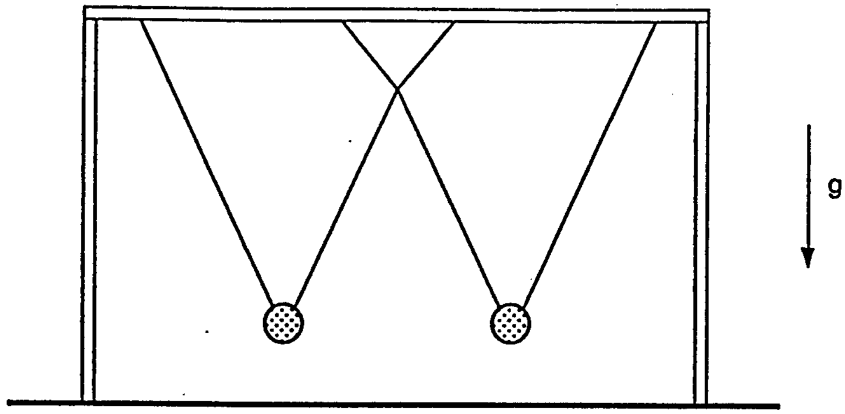
It is useful to first demonstrate motion purely in the y direction, and to note the relevant pendulum length, and then repeat this for motion purely in the x direction. These are the normal modes of the system. Next, if approximate circular motion is initiated, the motion is observed to quickly alternate between circular motion in the two directions. The motion appears to repeat, but this is not the case because the ratio of frequencies is an irrational number.

Finally, the connections of the string at the support are moved as far apart as possible, so that the angle of the V is nearly 180°. Circular motion is initiated in one direction, and the motion slowly alters, becoming a straight line, a circle in the opposite direction as the initial state, a straight line, and finally the initial state. The motion then approximately repeats. This is Lissajous motion for slightly different frequencies. Because the x and y motions have approximately the same frequency, the general motion is approximately elliptical. The phase between the x and y motions slowly changes, however, so that the motion slowly cycles through all possible elliptical motions.

The Lissajous pendulum offers a means of fooling observers (including physicists). A pendulum is made by inconspicuously looping the top of a string over a high structure such as a fluorescent light fixture. The pendulum mass is chosen to be nonmagnetic (e.g., aluminum), and is arranged to be near the floor. A magnet is placed just below the mass. Many people try to explain the behavior as due to electromagnetic induction! However, the behavior is simply Lissajous motion with slightly different frequencies (see above).

OS.6b Two coupled pendulums

Coupled pendulum apparatus. The masses oscillate perpendicular to the plane of the diagram.



More elaborate forms of this standard demonstration exist, but none matches the simplicity of this design, which is due to Robert Keolian at Pennsylvania State University. Two identical simple planar pendulums are constructed with string in a "V" shape. A knot is tied where the Vs overlap.

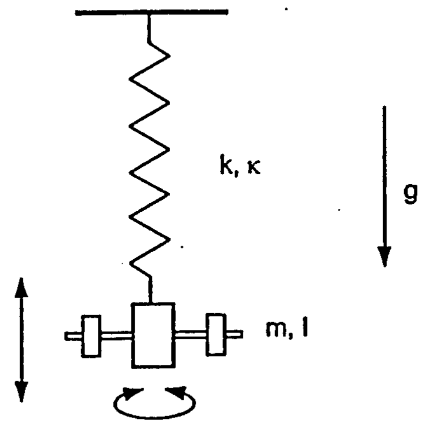
If one mass is set into oscillatory motion, while the other is initially at rest, the energy will "beat" back and forth between the two oscillators. The motion thus consists of two frequencies. The beat frequency is substantially smaller than the frequency of a single pendulum.

The two normal linear modes can easily be demonstrated. These modes, each of which has a pure frequency, consist of in-phase and out-of-phase motion with equal amplitudes. Note that the in-phase mode has the same frequency as a single pendulum (the coupling is effectively inoperative in this case). The knot is at rest in the out-of-phase mode, which results in a higher frequency because the effective length of a pendulum is shorter.

An interesting nonlinear effect can be demonstrated with this apparatus. The out-of-phase mode is stable at high amplitudes, whereas the in-phase mode is *unstable*. This is a general nonlinear instability which occurs for any system of coupled identical nonlinear oscillators that exhibit a breakdown of Hooke's law at finite amplitudes.[1] Pendulum oscillations *soften*; the frequency decreases with amplitude. To understand the instability, consider in-phase motion where pendulum A has a slightly greater amplitude than B. Because the frequency decreases with amplitude, A will lag behind B. This results in B doing a net amount of positive work on A over one cycle, so the amplitude of A will increase. This causes A to lag even farther behind B, which causes A to absorb even more energy from B, etc. The situation is thus analogous to a ball on a hill: The slightest displacement of the ball to one side causes the ball to move farther on that side, so the initial state is unstable.

[1]Bruce Denardo, John Earwood, and Vera Sazonova, "Parametric instability of two coupled nonlinear oscillators," *American Journal of Physics*, vol. 67, pp. 187-195 (1999).

OS.6c Wilberforce oscillator



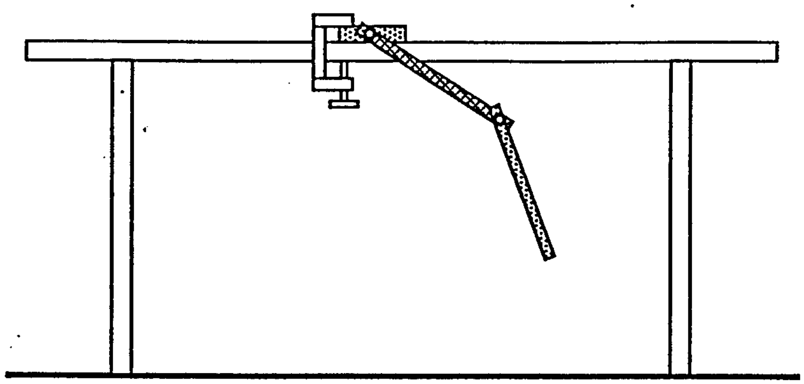
Wilberforce oscillator. The locations of the side masses are adjusted so that the frequencies of longitudinal and torsional oscillations are the same.

Longitudinal and torsional motions of a system of a mass suspended from a spring are carefully tuned to have the same frequency. The angular frequency of longitudinal oscillations is $\omega_L = (k/m)^{1/2}$, where k is the spring constant and m is the mass, and the angular frequency of torsional oscillations is $\omega_T = (\kappa/I)^{1/2}$, where κ the torsional constant and I is the moment of inertia. The mass of the spring has been neglected in both cases. Once the system has been built to yield roughly the same frequencies, the equality $\omega_L = \omega_T$ is achieved empirically by adjusting the locations of the side weights and thus the moment of inertia.

The system is initiated in either a pure longitudinal or torsional mode. Complete energy transfer back and forth between the two modes is dramatically observed. The situation is essentially the same as the coupled pendulum system (OS.6b). Here, however, the normal modes are not immediately obvious. The normal mode of lower frequency involves the spring uncoiling as it stretches, and coiling as it compresses. The normal mode of greater frequency involves the spring coiling as it stretches, and uncoiling as it compresses. An initial pure longitudinal or torsional excitation is an equal superposition of these two normal modes, which explains why the energy "beats" back and forth between the longitudinal and torsional modes.

Each normal mode can be demonstrated by carefully driving the top of the spring by hand to achieve an approximate pure-frequency response. An alternative is to raise or lower the mass and coil or uncoil the spring, and release the mass from rest. For a specific ratio of the linear and angular displacements, only one normal mode will be excited. The period of each normal mode can be measured with a stop clock (e.g., for 20 oscillations), and the values can be compared. The coupling is evidently weak because there is a long beat period in the normal demonstration above. Hence, the periods of the normal modes differ only slightly. It can be quantitatively confirmed that the beat frequency is the difference of the two normal mode frequencies.

OS.6d Chaotic double pendulum

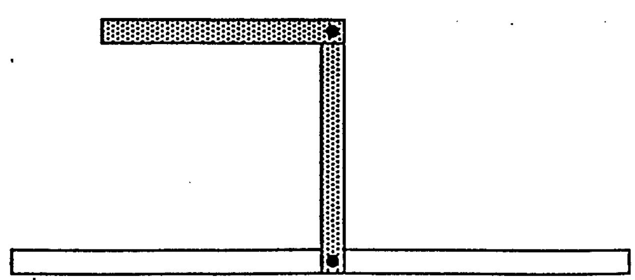


Double pendulum clamped to a table.

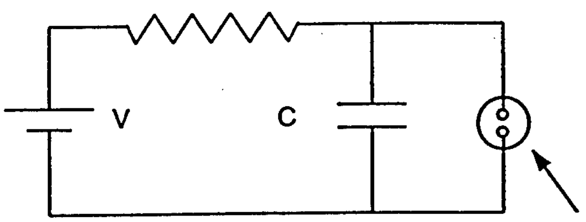
A double pendulum is constructed with aluminum bars and ball bearings such that the damping is small, and such that the arms can rotate through complete revolutions. The apparatus must be securely clamped to a sturdy table. By driving the top by hand to support small-amplitude motion, the demonstrator can exhibit the two linear normal modes (in-phase and out-of-phase monofrequency motions). By initiating motion in which the arms travel over the top, chaotic motion is observed.

Chaos is usually defined as extreme sensitivity to initial conditions. Specifically, if the initial conditions are slightly altered, the evolution of the system exponentially diverges from the original motion. This sensitivity to initial conditions can be demonstrated by releasing the system in the same position in successive trials. An appropriate initial condition that is easily repeatable is the upper pendulum inverted and the lower pendulum horizontal (refer to the bottom diagram). As other examples, the pendulums can be released from rest with the both pendulums inverted vertically, or with the upper pendulum inverted vertically and the lower pendulum normally vertical. The motion can be very different after several rotations. While demonstrating this, it is useful to point out some peculiar aspect of the motion during the first trial. It is unlikely that the same behavior will be observed during the second trial, even though the initial conditions were very nearly identical. An example of a peculiarity is the bottom arm of the pendulum rotating over-the-top in the same direction 4 or more times without reversing.

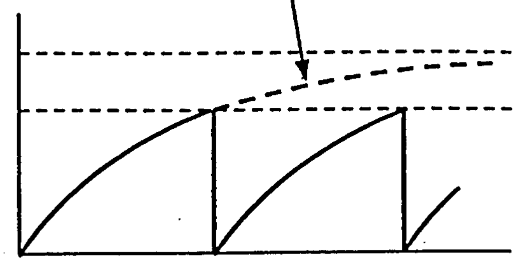
Initial condition useful to show sensitivity to initial conditions. The system is released from rest in the configuration shown.



OS.7 Relaxation oscillations of an RC and neon bulb circuit



Relaxation oscillator, where the breakdown voltage V_b must be less than the source voltage V.



Graph of voltage across the capacitor vs. time.

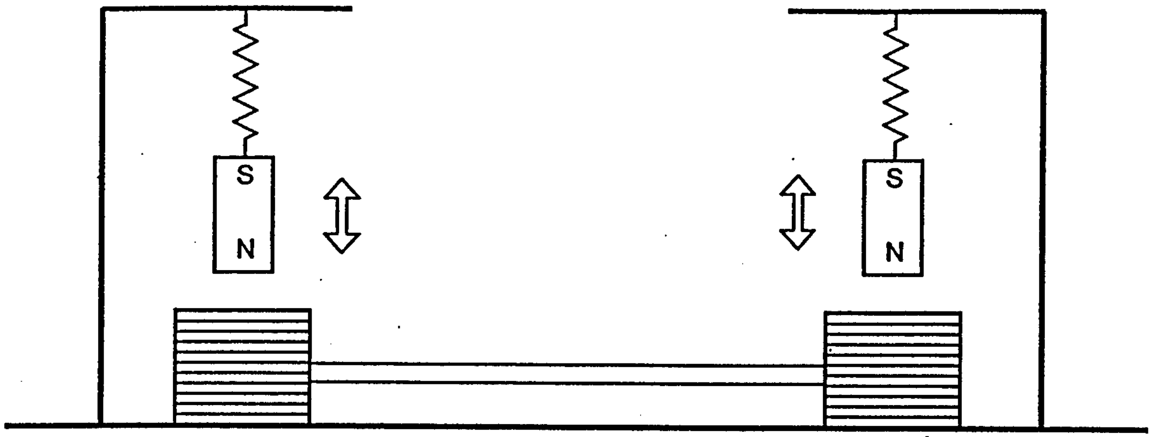
A relaxation oscillation is one in which the response steadily grows as a result of an external or internal drive until a maximum is reached, which triggers some mechanism that causes the response to decay. The process then repeats indefinitely. The decay of the response is typically substantially quicker than the growth. Relaxation oscillations occur in many systems. A simple example is the charging of a capacitor through a resistor, where a neon bulb across the capacitor causes a discharge before the capacitor is fully charged.

The neon bulb conducts and emits light at a breakdown voltage V_b. A variac can be employed to show this, although it should be noted that the light is dim when the peak voltage is at breakdown because the emf is alternating as opposed to dc. The breakdown voltage is less than 100 V.

A dc supply of approximately 100 V charges a capacitor C through a resistor R. The neon bulb is connected across the capacitor. Convenient values are R = 1.0 MΩ and C = 1.0 μF, which yield the time constant RC = 1.0 s. The charge on the capacitor, and thus the voltage across it, increase until breakdown of the bulb occurs. This produces a flash of light and shorts the capacitor, and the process then repeats. The bulb flashes with a period of roughly 1 s, although the period depends upon the relative values of V and V_b. For example, if V is only slightly greater than V_b, the period is substantially greater than RC. By considering the above graph, it not difficult to show that the exact flashing period is $RC \ln[V/(V - V_b)]$.

That the time constant is proportional to RC can be dramatically shown as follows. By adding a second 1.0 MΩ resistor in series with the original resistor, the resistance is doubled. This doubles the time constant, causing the original flashing period to double. The second resistor can then be added in parallel with the original resistor, which halves the resistance and thus halves the flashing period. A second 1.0 μF can be added in parallel to the original capacitor to double the capacitance, and thus double the original flashing period. The second capacitor can then be added in series with the original capacitor to halve it the capacitance, and thus halve the flashing period.

OS.8 Electromagnetically coupled oscillators



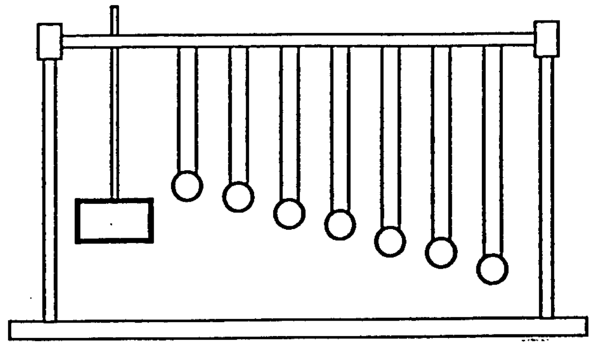
Two identical setups consist of a magnet suspended by a spring such that one pole of the magnet is near an end of a solenoid. The two solenoids are connected by wires, and are far apart. When one magnet is set into oscillation, the other begins to oscillate. In the final state, both magnets oscillate with equal amplitude either in-phase or out-of-phase. The phase is reversed if one magnet is inverted, or if the polarity of the connection is reversed.

When one magnet is set into oscillation, an oscillating emf is produced by Faraday induction in its solenoid. The resultant current causes the other solenoid to exert an oscillating magnetic force on the other magnet. In the final state, the emfs created by the oscillating magnets are out-of-phase and thus cancel. Hence, there is no current and the motion is only lightly damped. This is one normal mode of the system. In the other normal mode, the emfs are in-phase and thus add, so this mode is highly damped due to joule heating of the current. This explains why only one mode is eventually present.

The system of a single magnet and solenoid can be used to demonstrate electromagnetic damping. When the circuit is open (solenoid not connected to itself), the oscillations persist for a long time. When the circuit is closed, the oscillations quickly damp out. Due to the joule heating associated with the induced current, energy conservation requires that the oscillations dampen. The mechanism is essentially the same as in eddy current damping: the magnetic field caused by the current in the solenoid always impedes the motion of the magnet.

OS.9 Resonance pendulums apparatus

Resonance apparatus consisting of uncoupled flexible pendulums of different lengths suspended from a bar, which is oscillated by a heavy rigid pendulum.

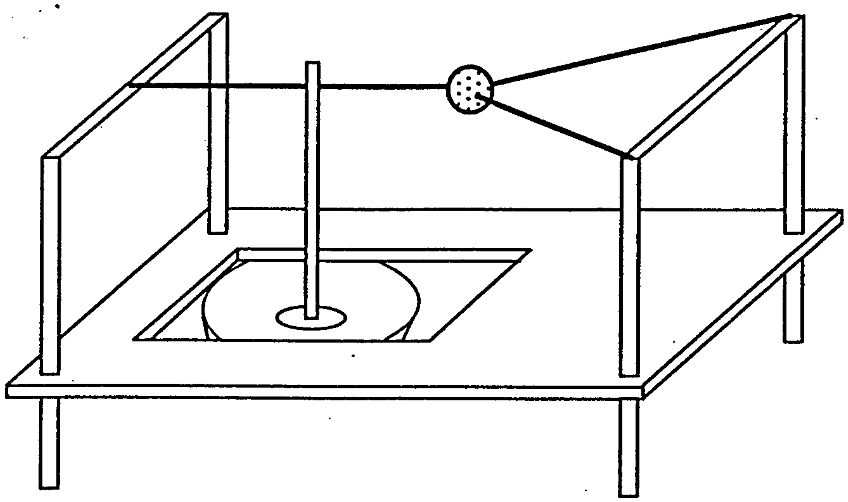


Seven light-weight flexible pendulums of successive lengths are attached to a horizontal bar that is driven by a heavy rigid pendulum attached near one end. The light pendulums are not coupled to each other. Common features of resonance can be demonstrated, including maximum amplitude on resonance and the phase of the response relative to the drive on and off resonance. The light pendulum having approximately the same length as the heavy pendulum has the greatest amplitude in the steady state. Furthermore, the phase of this light pendulum is observed to lag the phase of the heavy pendulum by 90°, and the phases of the other pendulums are observed to successively increase from 0° for the shortest to 180° for the longest. The apparatus thus effectively achieves a spatial display of the frequency response curve for a single oscillator as the drive frequency is slowly incremented in time.

Two fundamental problems occur in the construction of such an apparatus. First, unwanted longitudinal motion can occur due to the same parametric instability that can cause a vibrating string not to remain planar. Second, the transient time of the pendulums (roughly the time required to reach steady state motion) may not be short compared to the decay time of the heavy pendulum. Flexible strips of material for the pendulums are employed to reduce the longitudinal motion. The strips also increase the air drag and thus reduce the transient time. To further reduce the transient time, ping pong balls are used as the masses.

The center of the resonant ping pong ball lies about 1 cm below the center of the heavy mass. The effect is not due to the extended nature of the masses or to dissipation, both of which can be shown to be negligible. There are three contributing causes: the natural frequency of a light pendulum is increased due to the stiffness and mass of a strip, and the natural frequency of a light flexible pendulum corresponds to the length from the *bottom* of the horizontal bar whereas the frequency of the heavy rigid pendulum corresponds to the length from the *center*. Surprisingly, this third effect nearly completely accounts for the 1 cm distance, even though the strip mass effect alone is calculated to produce a 1/2 cm shift. There is an additional effect here that roughly cancels the effect of the mass and stiffness of the strips: the "hydrodynamic" mass, which is due to motion of the ping pong ball and strip through the air. This added inertia decreases the frequency of a pendulum.

OS.10 Driven nonlinear oscillator



A mass-and-rubber band oscillator is driven vertically by a loudspeaker. (The function generator and amplifier are not shown.) Nonlinear shifting of the resonance, including hysteresis, are demonstrated.

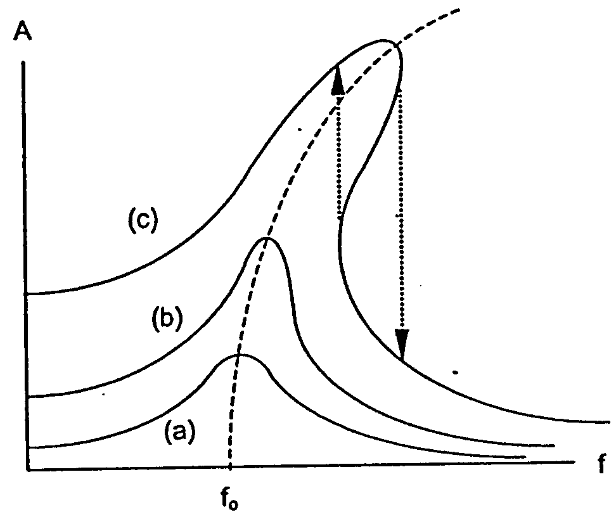
This demonstration was created by Robert Keolian (Graduate Program in Acoustics, Pennsylvania State University). Rubber bands are attached to supports and a wooden ball (see diagram above). A rod attached to a loudspeaker drives vertical oscillations of the ball. An 8-inch diameter 8-Ω loudspeaker is appropriate. The "V" arrangement breaks the degeneracy between the vertical mode and transverse horizontal mode. Otherwise, the vertical mode will readily parametrically excite this horizontal mode, in the same way that a string vibrating in a plane can develop motion perpendicular to the plane.

In the first part of the demonstration, the low-amplitude (approximately linear) response of the system is shown. A suitable voltage to the loudspeaker is 0.5 V. The drive frequency is finely (0.01 Hz) and slowly increased so that the motion is approximately steady state. Resonance occurs at 11.55 Hz for one arrangement, and the response is observed to be roughly symmetric about resonance.

Next, the voltage is increased roughly an order of magnitude (4.0 V is suitable). The drive frequency is again slowly and finely increased, but now the maximum response amplitude is large and occurs at 12.05 Hz, after which the amplitude discontinuously decreases by a substantial amount. Further increase in the drive frequency causes a continuous decrease in amplitude. The drive frequency is now decreased, and the amplitude continuously increases until 11.92 Hz, at which the amplitude discontinuously increases. Further decrease in the drive frequency causes the amplitude to continuously decrease. The system thus exhibits hysteresis when the response amplitude is sufficiently large. For a drive frequency roughly midway between the jump frequencies (for example, 12.00 Hz), there are thus *two* stable response states, which can be demonstrated by appropriately adjusting the initial conditions by hand.

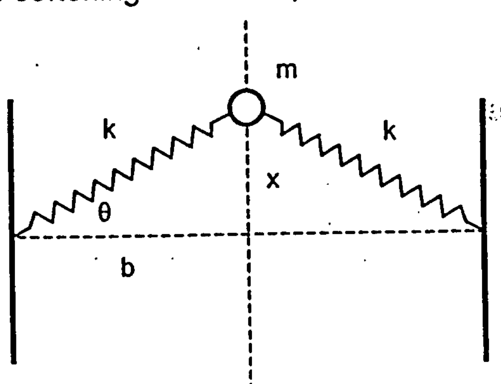
This behavior occurs because Hooke's law breaks down for finite amplitudes of oscillation here (see explanation below). At small response amplitudes, the motion is approximately linear, and the frequency-response curve (steady-state amplitude vs. drive frequency for a fixed drive amplitude) appears as in (a) in the graph below. For greater response amplitudes, a shift of the resonance frequency becomes apparent. The frequency-response curve conforms to the "backbone" curve for free oscillations. At sufficiently large response amplitudes, the frequency response curve becomes multivalued as in (c) in the graph, giving rise to hysteresis. In the multivalued region, the largest and smallest amplitudes are stable, whereas the intermediate amplitude is unstable.

Frequency-response curves for a driven nonlinear oscillator. The response is approximately linear in (a), a shift of the resonance frequency is apparent in (b), and hysteresis (dotted lines) occurs in (c). The dashed "backbone" curve corresponds to free oscillations. The linear natural frequency is f_o.



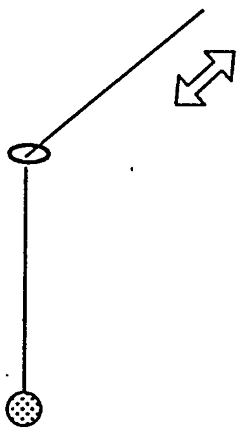
To understand the nonlinearity, consider the simplified oscillator shown in the figure below. Let k be the spring constant of each spring, and a be the unstretched length. Each spring is stretched to a distance $b \geq a$ in equilibrium. When the mass has displacement x, the net restoring force is $F = 2F_o\sin(\theta)$, where the force due to each spring is $F_o = k[(b^2 + x^2)^{1/2} - a]$ and where $\sin(\theta) = x/(b^2 + x^2)^{1/2}$. Hence, $F = 2kx[1 - a/(b^2 + x^2)^{1/2}]$. For $x/b << 1$, the approximation $(b^2 + x^2)^{-1/2} = b^{-1}(1 - x^2/2b^2)$ yields $F = 2k(1 - a/b)x + (ka/b^3)x^3$. This shows that the oscillations *harden* (the stiffness increases with amplitude). In the calculation, note that the increase in tension causes hardening while the geometry [i.e., $\sin(\theta)$] causes softening. However, the net result is hardening.

Transverse displacement x of the mass of a mass-and-spring oscillator. For finite amplitudes, the restoring force is not linear in x.



OS.11 Parametrically driven pendulum

Pendulum parametrically driven by hand such that the length of the pendulum is modulated.

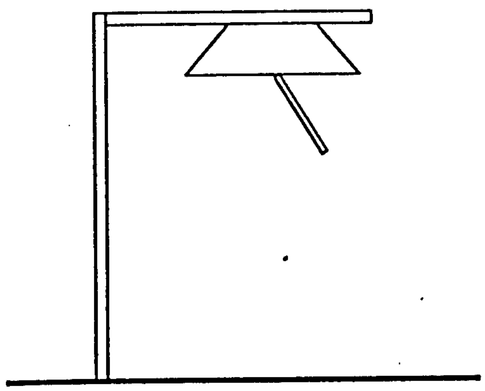


Pendulum parametrically driven by hand such that the acceleration due to gravity is modulated.

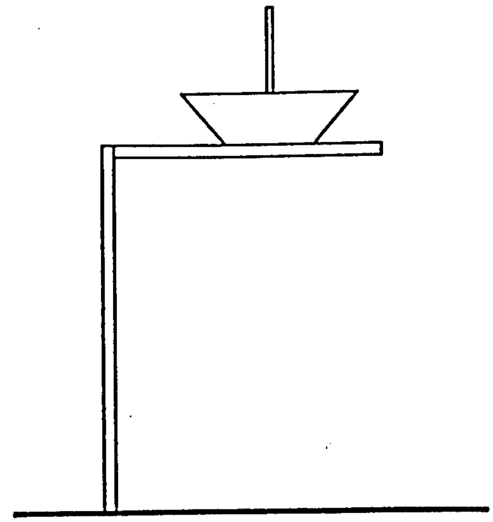
It is instructive to begin by demonstrating a pendulum that is parametrically driven by hand. A weight on a string can serve as the pendulum. Two means of parametric excitation can be shown. In the first, an eyelet around the string is formed with the index finger and thumb of one hand, which is held fixed. The other hand alternately shortens and lengthens the portion of the string below the eyelet. If this drive frequency is approximately twice the natural frequency of the ambient pendulum, and if the drive amplitude is sufficiently large, parametric excitation of the pendulum will occur. This is referred to as "parametric" because the excitation results from the modulation of a parameter (in this case, the length of the pendulum) upon which the natural frequency depends.

Second, the upper end of the pendulum is held with one hand, which thus serves as the support for the pendulum. The hand is oscillated vertically. For values of the drive parameters as in the previous case, parametric excitation will occur. In the noninertial frame of reference in which the support is at rest, the acceleration due to gravity is effectively modulated according to Einstein's principle of equivalence. (For example, a person in an elevator that is oscillating up and down would alternately feel heavier and lighter, respectively.) Hence, the relevant parameter in this case is the acceleration due to gravity.

A controlled parametrically driven pendulum can be constructed by hinging a hollow plastic rod to the cone of a loudspeaker, which is driven by a function generator and a power amplifier. This demonstration was created by Robert Keolian (Pennsylvania State University). A 4.5-inch rod and an 8-inch diameter 8-Ω loudspeaker are appropriate (see diagram below). In the first demonstration with this apparatus, the loudspeaker is driven at 3.70 Hz and 3.0 V. The pendulum is given a push, and parametric excitation occurs. After steady state motion obtains, the drive frequency is decreased in increments of 0.01 Hz, and the response amplitude is observed to increase. At approximately 3.30 Hz, the response dies to zero. The drive frequency is now increased, while the pendulum is given slight pushes between increments, but the response remains zero. At 3.50 Hz, the motion returns. This shows the hysteresis that is inherent in parametric excitation (see graph below).

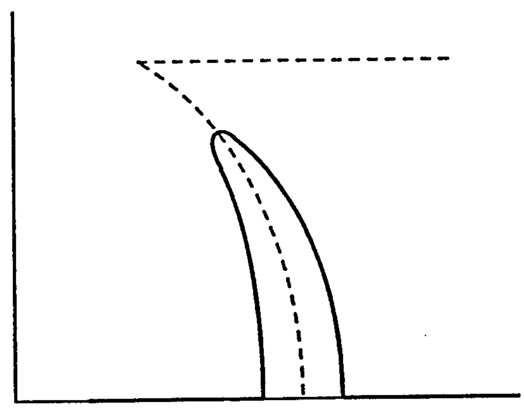


Hinged pendulum connected to a loudspeaker, which acts as a parametric drive. (The function generator and amplifier are not shown.)



The loudspeaker is now inverted, so that it parametrically drives an inverted pendulum.

In the next demonstration, the drive frequency is again initially 3.70 Hz, but now the drive amplitude is doubled to 6.0 V. The frequency is slowly reduced and the response amplitude increases until the rod touches the rim of the loudspeaker. Now the drive frequency is *increased* to roughly 4.0 Hz. The contact with the rim has reversed the original gravitational softening of the oscillations (natural frequency decreasing with amplitude), causing the oscillations to now harden.

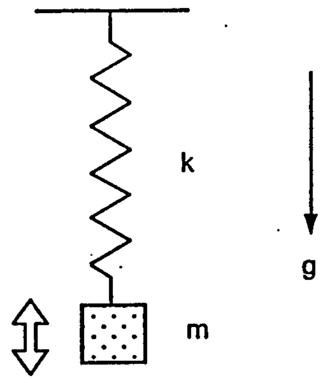


Steady-state frequency-response curve for a parametrically driven pendulum. The dashed line represents free (undriven) oscillations. The lower part of the curve corresponds to unstable motion.

In the final demonstration, the loudspeaker is inverted. The vertical orientation of the pendulum is now unstable. However, by parametrically driving the pendulum, it is possible to stabilize it. Suitable drive parameters are 25.0 Hz and 6.5 V.

OS.12 Parametric instability

A mass connected to a spring initially undergoes vertical oscillations, but then develops transverse motion.

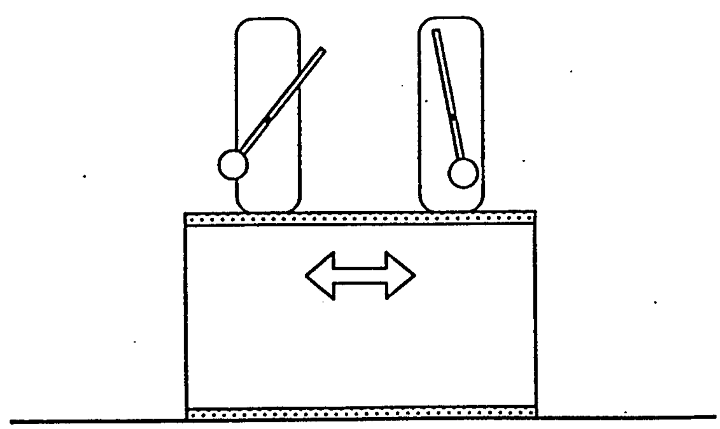


A mass connected to a spring is made to undergo vertical oscillations. By adjusting the spring constant or mass, or frequently not adjusting either, it is observed that vertical oscillations are unstable; specifically, that pendulum motion unavoidably develops. The motion "beats" between mass-spring and pendulum oscillations. This is often considered to be a nuisance, but the phenomenon is interesting. It occurs most readily when the mass-spring frequency $f = (k/m)^{1/2}/2\pi$ is *twice* the pendulum frequency $(g/L)^{1/2}/2\pi$, where L is the ambient length of the spring-mass system. This is a *parametric instability*, because the mass-spring mode is driving the pendulum mode by modulating the length of the pendulum. When the mass-spring frequency is only roughly equal to twice the pendulum frequency, there is an amplitude threshold for the instability to occur. This instability has caused ships to capsize. Waves can cause a ship to bob up and down. If the frequency of these oscillations is roughly twice the frequency of the side-to-side rocking motion of the ship, then the rocking motion can grow to the extent that the ship capsizes.

Two coupled pendulums also exhibit a parametric instability. Refer to OS.6b.

OS.13a Mode locking of two coupled metronomes

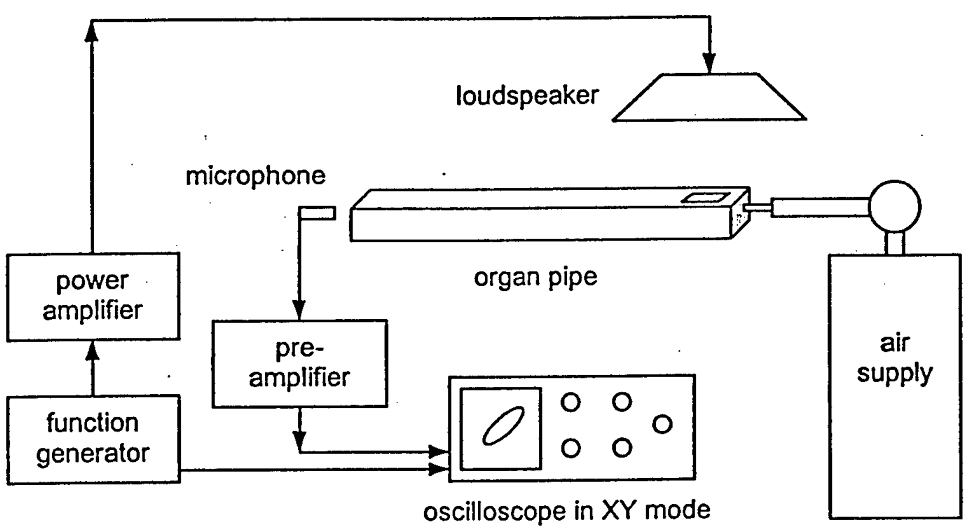
Two mechanical metronomes with slightly different frequencies are placed on a nonrigid platform. As a result of the coupling and the drives, the metronomes eventually lock to the same frequency.



Two metronomes with driven pendulums are set at low frequencies that are slightly different. First, the metronomes are placed on a rigid table, so that their coupling is negligible. The frequencies are seen and heard to be different because the phase between the metronomes slowly cycles through 0°, 90°, 180°, and –90°. Next, the metronomes are placed on a platform with hacksaw-blade supports such that the blades flex in response to the motion of the pendulums. The metronomes now eventually mode lock, so that their frequencies are the same, which can be both seen and heard. The metronomes tend to lock approximately out of phase, although approximate in-phase locking is possible.

Another means of coupling the metronomes so that mode locking can occur is to attach a strong small neodymium-iron-boron magnet to each pendulum bob. The metronomes are then arranged to face each other at a distance where the coupling is negligible. One metronome is then slowly pushed toward the other. At a critical distance, the metronomes will mode lock. This can be done such that the magnets either attract or repel.

OS.13b Mode locking of an organ pipe to a loudspeaker



An organ pipe is driven by an air supply, and produces a tone. When a loudspeaker driven at a slightly different frequency is brought near the organ pipe, the tone can lock to the frequency of the loudspeaker.

This demonstration was created by Robert Keolian (Graduate Program in Acoustics, Pennsylvania State University). An air supply is fed to an organ pipe, which produces a tone. A loudspeaker is driven at a slightly different frequency, as evidenced by beats that are heard. This is also shown on an oscilloscope by a steadily rotating Lissajous pattern. When the loudspeaker is brought near the organ pipe, the rotation ceases to be steady. As the loudspeaker is brought even nearer to the organ pipe, there is a critical point at which only a single frequency is heard and the Lissajous pattern ceases to rotate. The organ pipe has then become mode-locked to the loudspeaker.

Mechanical Waves

1. Speed of sound in solids: collision balls and glider collisions of mass ratio 3

2. Aluminum rods: excite and reveal overtones. Doppler effect

3. Resonant acoustic breaking of a glass

4. Waves in nonuniform systems: standing waves on torsional wave machine impedance matcher, and variably nonuniform resonators

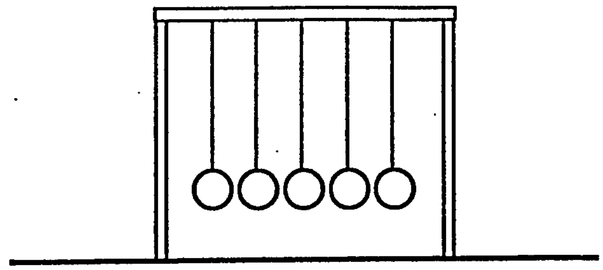
5. Nonradiating wave source: hot wire in magnetic field

6. Dispersion: magnet lattice

7. Solitons: pendulum lattice and gravity waves

MW.1a Collision ball apparatus

The collision ball apparatus is a familiar toy. When a ball at one end is pulled back and released, the ball at the other end rises to nearly the same initial height. Why does only one ball depart from the end?



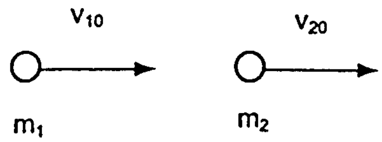
A ball at one end of the apparatus is pulled back and released from rest at some height. The ball comes to rest after the collision, and the ball at the other end emerges to nearly the same height, showing that the collisions are elastic. For good apparatus, the process repeats for a substantial number of cycles. Why does only one ball emerge? For example, why cannot two balls emerge together and rise to a lesser height (so that energy is conserved)? The answer is that the former is the only way in which both momentum and energy are conserved. The behavior can be explained by imagining that the balls are separated by a small distance. (For the importance of this, see below.) What occurs is then a series of binary equal-mass elastic collisions where each target mass is initially at rest. Hence, the initial ball comes to rest and the final ball is ejected with a speed equal to the initial speed.

Two balls at one end can be pulled back together, and then released from rest. The same can be done for three and four balls. If the incident balls as well as the other balls are imagined to be separated by a small distance, the idea of individual equal-mass collisions can be used to explain the behavior in these cases. Another demonstration is to pull back two balls on one end and one ball on the other, and simultaneously release the balls from rest. Other possible demonstrations of this type can also be done.

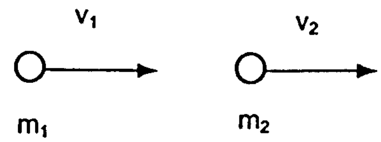
It is interesting to use a piece of double-sided tape to join two balls at one end. This can be done either openly or surreptitiously. The latter "trick" is appropriate because many students are aware that the number of incoming balls equals the number of outgoing balls. When a single ball on the other end is pulled back and released, what happens? The double-ball combination rises to a lesser height than the incident ball, other balls also exit, and the incident ball rebounds.

A small gap between the balls is not only convenient in explaining the behavior, but necessary to achieve good results with the apparatus. The problem is that collisions in general are more complicated if they are not binary (refer to MW.1b). One might think that the collisions would be binary even if the balls touch in equilibrium, due to the finiteness of the speed of sound in the balls. However, for this to obtain, the duration of a collision between two balls must be shorter than the transit time of the sound wave. This is evidently not the case, because a small separation distance is observed to lead to much better results.

MW.1b Head-on collision with bodies of mass ratio 3



Initial state of two bodies.



Final state after a one-dimensional collision.

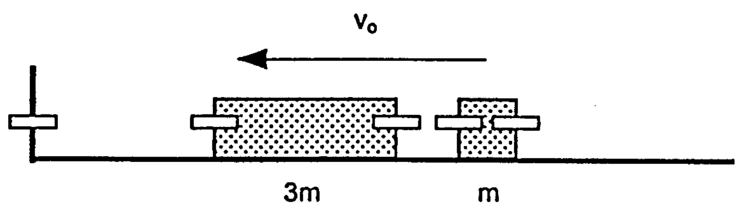
For a one-dimensional elastic collision of two bodies with masses m_1 and m_2 and respective initial velocities v_{10} and v_{20}, the final velocities are

$$v_1 = \frac{(m_1 - m_2)v_{10} + 2m_2v_{20}}{m_1 + m_2} \quad \text{and} \quad v_2 = \frac{2m_1v_{10} - (m_1 - m_2)v_{20}}{m_1 + m_2}.$$

Consider a head-on collision in which the masses have the same initial speed v_0. Setting $v_{10} = -v_{20} = v_0$ in the equations yields $v_1 = (m_1 - 3m_2)v_0/(m_1 + m_2)$ and $v_2 = (3m_1 - m_2)v_0/(m_1 + m_2)$. Hence, if $m_1 = 3m_2$ a particularly simple final state occurs: $v_1 = 0$ and $v_2 = 2v_0$.

Qualitatively, for a collision between mass m_1 and m_2 in which the velocities are equal and opposite, m_1 must continue to move in the same direction if $m_1 >> m_2$. For $m_1 = m_2$, m_1 will move in the opposite direction with the same speed. Hence, there must be a critical value of m_1 between m_2 and a value much greater than m_2 such that m_1 will be at rest after the collision. The above theory shows that this value is $3m_2$.

Two gliders move with the same velocity toward the end of an air track. The larger glider is at rest after the collisions.



A one-dimensional collision in which two bodies have equal and opposite initial velocities can be arranged with gliders on an air track by giving the gliders the same velocity v_o and allowing the front glider to collide elastically with the end of the track. The gliders and end of the track should have spring bumpers. After the first collision, 3m is moving with velocity v_o toward m. After the second collision, 3m is predicted to be at *rest* and m moves with velocity $2v_o$ opposite to the initial direction. The result is not undramatic. It is interesting that *binary* collisions must be made to occur here. If m starts to collide with 3m before the collision between 3m and the boundary is completed, 3m will *not* be at rest in the final state, but will be moving away from the boundary. This can be demonstrated by carefully arranging the two gliders to remain in contact or near contact as they travel toward the end of the track.

MW.2a Longitudinal vibration rods

Longitudinal vibration rod for *odd* harmonics, with markings at nodes of the 1st, 3rd, and 9th longitudinal modes. "C" denotes the rod's center, which is a node for all odd-numbered modes including the fundamental (1st). The L3 nodes are also L9 nodes.

Longitudinal vibration rod for *even* harmonics, with markings at nodes of the 2nd, 6th, and 10th longitudinal modes. The L2 nodes are also L6 and L10 nodes.

Solid aluminum rods with diameter 1/2 inch and length 6 feet are employed to demonstrate longitudinal standing waves. A rod with the same dimensions is used in the next demonstration (MW.2b) to show transverse standing waves. Circumferential notches mark the nodes of various free-free modes. Stamped near each notch on the rod are the designations Ln, where "L" refers to "longitudinal" and n is the number of the corresponding mode. The nodal positions are the zeros of the displacement function $\cos(n\pi x/L)$, where L is the length of the rod, and are listed in the two tables below. In the tables, $c = (Y/\rho)^{1/2}$ is the speed of longitudinal waves, where Y is the Young's modulus and ρ is the density. The nodal positions are listed to the nearest 0.1%. For demonstrations, this yields an acceptable precision of ±1 mm for a rod of length 6 feet.

Frequencies and nodal locations of odd harmonic longitudinal modes of a free-free rod.

mode	frequency ($f_1 = c/2L$)	nodal positions measured from one end (percentage of length of rod)
L1	f_1	50.0
L3	$3 f_1$	16.7, 50.0, 83.3
L9	$9 f_1$	5.6, 16.7, 27.8, 38.9, 50.0, 61.1, 72.2, 83.3, 94.4

Frequencies and nodal locations of even harmonic longitudinal modes of a free-free rod.

mode	frequency ($f_1 = c/2L$)	nodal positions measured from one end (percentage of length of rod)
L2	$2 f_1$	25.0, 75.0
L6	$6 f_1$	8.3, 25.0, 41.7, 58.3, 75.0, 91.7
L10	$10 f_1$	5.0, 15.0, 25.0, 35.0, 45.0, 55.0, 65.0, 75.0, 85.0, 95.0

A rod is held between a thumb and forefinger at a desired node. Longitudinal sound modes are generated by pulling the rod between the other thumb and forefinger that have been rubbed with rosin. Another means of exciting sound is to strike the end of the rod against the floor, although the rosin method is much more effective. Each rod is marked at points where it can then be held to eliminate the primary mode and thus show that higher modes are present.

Rosin is essential for good results. It greatly enhances the stick-slip action of the thumb and forefinger rubbing the rod, and thus leads to much larger response amplitudes. This same action occurs for the motion of a bow on a stringed instrument. (Indeed, the rosin is easily obtained at musical instrument stores.) To properly coat the thumb and forefinger, small pieces of rosin should be rubbed between the thumb and forefinger. Small pieces can be broken off the rosin by lightly tapping an edge of it on a hard surface.

The sound can be very loud, even to the point of being painful to the ear. Due to the loudness and the slow decay time, many students think that the rod is made of some special material, or that there is a hidden source of energy. It should be pointed out that this is not the case; the quality factors of the various acoustic modes of metal rods and plates are typically very high.

The odd-mode rod is held at the center C, so that the 1st mode is primarily excited by the rubbing. To show that the 3rd mode is also present, the demonstrator then holds the rod at an L3 node other than the center. This very quickly damps the 1st mode, and the pitch jumps by a factor of 3 (an octave and a fifth). The next mode present is the 9th, which is demonstrated by holding the rod at an L9 node different from the L3 nodes. The pitch again jumps by a factor of 3.

The 2nd mode can be demonstrated by holding the even-mode rod at either L2 node and rubbing the rod. It is useful to demonstrate the fundamental immediately prior to this, so that the fact that the second mode is an octave higher (twice the frequency) can be heard. Modes 6, 10, 14, ..., are also excited, but not modes 4, 8, 12, ... In fact, this is easier to see than in the corresponding odd-numbered mode case, because both the ends and the center are antinodes for even-numbered modes (refer to the standing wave diagrams). That the 6th mode is present can be demonstrated similarly as in the case of the odd-numbered modes: After exciting the rod by holding it at L2, the demonstrator holds it either halfway between L2 and C or halfway between L2 and the nearer end.

The use of a microphone and oscilloscope allows the frequencies of the modes to be measured and compared to theoretical values in the above tables.

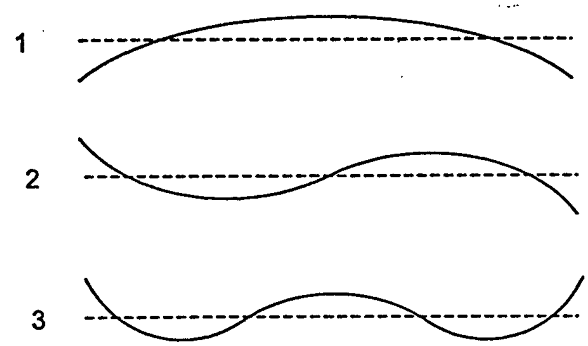
After demonstrating longitudinal and transverse modes of a rod, the instructor can ask students if another type of wave can propagate in a rod. This makes contact with the torsional wave machine (MW.4a).

MW.2b Transverse vibration rod

Transverse vibration rod, with markings at nodes of the first three transverse modes. "C" denotes the rod's center, which is a T2 node.

A solid aluminum rod of the same size as the longitudinal vibration rod (1/2-inch diameter and 6-foot length; MW.2a) is used to demonstrate transverse (flexural) standing waves, although the low modes typically do not have sufficient amplitude to be heard.

First three transverse standing wave modes of a free-free rod.



As in the longitudinal rod, circumferential notches mark the nodes of various free-free modes, in this case the first three. The designation T1, T2, or T3 of the corresponding mode is stamped near each notch on the rod. Due to the presence of hyperbolic functions in addition to circular trigonometric functions describing transverse standing waves on a bar, the frequencies and nodal positions must be numerically solved, although the spectrum quickly approaches a simple relationship. The results are displayed in the table below. In the table, L is the length of the rod, $c = (Y/\rho)^{1/2}$ is the speed of longitudinal waves, where Y is the Young's modulus and ρ is the density, and $\kappa = a/2$ is the radius of gyration, where a is the radius of the rod. For demonstrations, nodal positions to the nearest 0.1% yield an acceptable precision of ±1 mm for a rod of length 6 feet.

Frequencies and nodal locations of the first three transverse modes of a free-free rod.[1]

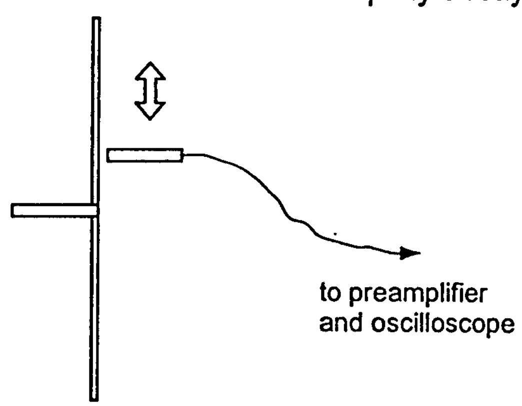
mode	frequency ($f_o = \pi\kappa c/8L^2$)	nodal positions measured from one end (percentage of length of rod)
T1	$3.0112^2 f_o$	22.4, 77.6
T2	$5.0000^2 f_o$	13.2, 50.0, 86.8
T3	$7.0000^2 f_o$	9.4, 35.6, 64.6, 90.6

The rod is initially supported horizontally by hand at the two nodes of the fundamental (T1). The rod should rest on the straightened index finger of each hand, with the remaining fingers straightened immediately below for support. For control, the thumbs should be lightly rested on the top of the rod. The rod is then quickly lowered as the knee of a leg is lifted in order to strike the rod at the center. For a moderately large impulse by the knee, the fundamental transverse mode is clearly visible for more than several seconds.

The importance of supporting the rod at the nodes of the desired mode can be demonstrated by repeating the above with the hands at arbitrary asymmetrical positions. In this case, the response of the rod quickly dampens. Excitation of the third (T3) mode can be accomplished as in the case of the fundamental, but with the hands at a pair of diametrically opposed T3 nodal positions. This can then be repeated with the other pair of T3 nodal positions to indicate the existence of a total of four nodes.

The second (T2) mode can be demonstrated as follows. A microphone is connected to a voltage amplifier whose output is connected to an oscilloscope. The rod is held vertically between a thumb and forefinger at the center of the rod, and then the palm of the other hand is used to impart an impulse at roughly the middle of either half of the rod. After a delay of several seconds, during which the higher modes decay, the microphone is scanned along either half of the rod from the center to the end. The microphone should be near the rod but not touch it. The signal due to the sound emitted by the vibrations is displayed on the oscilloscope. The node at the center, the antinode, the node between the center and the end, and the motion of the end of the rod are all clearly visible. The demonstration can also be done by imparting a large impulse to the rod, and immediately scanning with the microphone. Distortions of the waveform, which are due to the presence of higher standing wave modes, are then observed on the oscilloscope. The distortions are observed to rapidly decay.

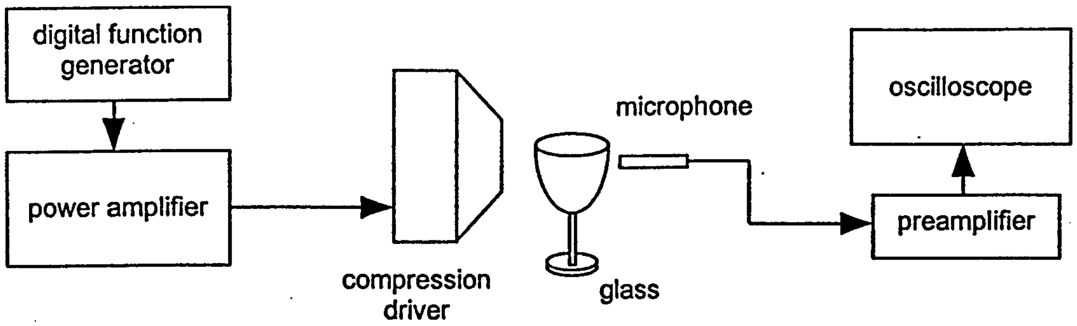
Demonstration of the second flexural mode. The rod is held at its center, and is transversely struck. A microphone is scanned by hand along the rod to show the response.



The use of a microphone and oscilloscope or spectrum analyzer allows the frequencies of the modes to be measured and compared to theoretical values in the above table. For $Y = 7.1 \times 10^{10}$ Pa, $\rho = 2.7 \times 10^3$ kg/m^3, we find $c = (Y/\rho)^{1/2} = 5.13 \times 10^3$ m/s. For $a = 0.25$ inch $= 6.35 \times 10^{-3}$ m and $L = 6.0$ feet $= 1.83$ m, we then find $f_o = 1.91$ Hz. Hence, the frequencies of the first three modes are $f_1 = 17$ Hz, $f_2 = 48$ Hz, and $f_3 = 94$ Hz.

[1]Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), pp. 75-76.

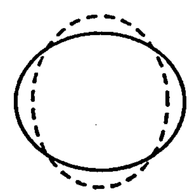
MW.3 Resonant acoustic breaking of a glass



Arrangement for breaking a glass by acoustic resonant excitation of a mode.

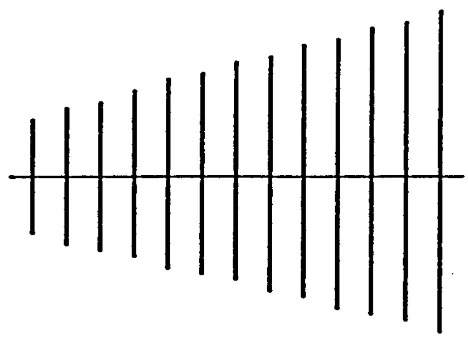
A wine glass, glass beaker (e.g., from 100 to 500 milliliters), or other type of thin glass is broken by acoustically driving the fundamental mode of the glass (see diagram below). The glass can be supported by machine nuts, watch glass, or some other hard object or objects that contact the glass over a small area, in order to minimize the damping of the mode of the glass (improve the *quality factor*). If, however, the glass "walks" while subjected to the high-amplitude sound, then the quality factor should be improved by using double-stick tape on the bottom of the glass to prevent it from moving.

Top view of the fundamental or "bell" mode of a glass.



A procedure for the demonstration is as follows. The glass is tapped on its rim, and a pitch is heard. If the oscillations decay too quickly for a definite pitch to be heard, then the support for the glass should be improved such that the oscillations persist longer. A frequency near that of the pitch is selected on a digital function generator. For a moderate drive amplitude, the frequency is then varied while the response of the glass is monitored on an oscilloscope connected to a microphone and a preamplifier. The resonance frequency is found by maximizing the response amplitude. If this maximum is only slightly greater than amplitudes for frequencies far from resonance, then the quality factor needs to be improved. The resonance frequency should be determined with a precision of roughly one part in 10^4, which is the reason that a digital function generator should be used. Hunting for the resonance is instructive, and should be done while the audience observes. The microphone is removed, the demonstrator inserts ear plugs and puts on safety goggles, and the audience is instructed to put their fingers in their ears. The drive amplitude is then increased until the glass breaks. If the increase is rapid, the glass typically splits roughly in half either along the plane that bisects the driver or along the perpendicular plane, consistent with points of maximum strain of the mode. If the increase is slow, the glass typically fractures into many pieces.

MW.4a Standing waves in a nonuniform medium



Top view of torsional wave apparatus with a nonuniform wave speed. (The speed decreases from left to right due to increased inertia.) The supports are not shown.

This demonstration utilizes the Pasco torsional wave apparatus in which the length of successive cross rods steadily changes from a short to a long value (see diagram above). The apparatus is marketed as an impedance matcher for two uniform torsional wave apparatus whose cross rod lengths equal the short and long values. One end of the nonuniform apparatus is driven with the included shaker at a frequency corresponding to a standing wave resonance. For example, a clearly visible standing wave with 5 half-wavelengths occurs if one end is fixed and the other is driven at 3.25 Hz and 5.0 V rms. It is observed that the wavelength decreases and the amplitude increases from left to right in the above diagram. If the driver is on the right, one might think that the amplitude is greater to the right as a result of attenuation. That this is not the case can be shown by driving on the left, which yields the same variation in amplitude.

To understand the variation in wavelength and amplitude, we consider traveling waves rather than standing waves. This is entirely general because a standing wave is the superposition of two counter-traveling waves even in a nonuniform medium, as long as a traveling wave changes adiabatically (gradual nonuniformity so that no reflections occur). It is instructive to first consider the simpler case of transverse waves on a string whose density increases from left to right. For a wave traveling along the string, the frequency f must be constant; otherwise, a steady state would not occur. If the wave travels from left to right, the wavelength λ decreases because the wave speed $c = \lambda f$ decreases due to the increasing inertia. Because $c \propto \rho^{-1/2}$, where ρ is the mass per unit length, the wavelength varies as $\lambda \propto \rho^{-1/2}$.

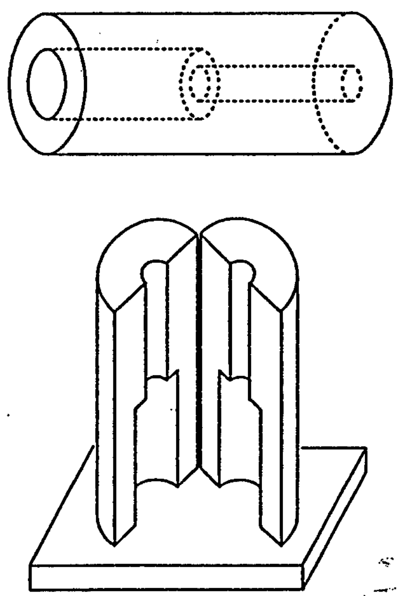
The change in amplitude can be understood and quantified in the case where the wave changes adiabatically. The average energy per unit length of a traveling wave is twice the average kinetic energy per unit length: $\rho\omega^2A^2/2$, where A is the displacement amplitude and $\omega = 2\pi f$. The energy in a wavelength $\lambda = 2\pi c/\omega$ is thus $\pi\rho c\omega A^2$, which must remain constant for adiabatic propagation. Because $c \propto \rho^{-1/2}$, we find that the energy per wavelength is proportional to $\rho^{1/2}A^2$. An adiabatic invariant is thus ρA^4, so the amplitude varies as $A \propto \rho^{-1/4}$. Because ρ increases to the right, the amplitude A must decrease, in *contrast* to what is observed for the torsional wave apparatus.

For the torsional wave apparatus, the wave speed is $c \propto I^{-1/2}$, where I is the moment of inertia per unit length. Because $I = \rho L^2/12$, where L is the rod length, and ρ

$\propto$ L, we find that $I \propto \rho^3$ and thus $c \propto \rho^{-3/2}$. The wavelength thus varies as $\lambda \propto \rho^{-3/2}$ (in contrast to $\rho^{-1/2}$ for transverse waves on a string). The average energy per unit length is $I\Omega^2/2$, where Ω is the angular frequency amplitude of a rod. If A is the displacement amplitude of the end of a rod, the angular frequency is the velocity amplitude ωA divided by the distance L/2: $\Omega = 2\omega A/L$. The energy per unit length is thus $2I\omega^2A^2/L^2$, which is proportional to ρA^2, and the energy in a wavelength is proportional to $\rho^{-1/2}A^2$. An adiabatic invariant is thus $\rho^{-1}A^4$, and the amplitude varies as $A \propto \rho^{1/4}$ (in contrast to $\rho^{-1/4}$ for transverse waves on a string). This predicts that the amplitude *increases* from left to right, which agrees with the observation.

The Pasco nonuniform torsional wave apparatus has minimum and maximum rod lengths that differ by a factor of 2.0. Ratios of measurements of the half-wavelengths and amplitudes at either end are in good agreement with the predicted wavelength ratio $2^{-3/2}$ and amplitude ratio $2^{1/4}$.

MW.4b Variably nonuniform resonators



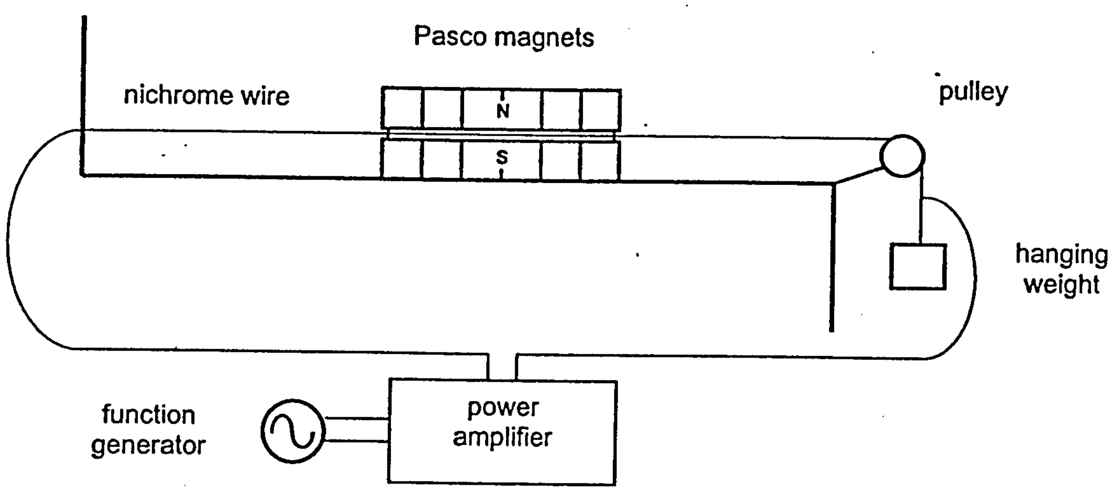
Acoustic resonators with a step discontinuity in cross-sectional area. In the first case, the frequency of the fundamental mode is heard when either end is slapped by hand. The two frequencies are different. In the second case, the two halves are clapped together to sound the fundamental. Three different frequencies are heard depending upon the orientations of the halves.

Acrylic rods have holes of different diameter bored from either end and meet in the middle. One rod is a single piece, while the other is cut in half along its length. An end of the first is slapped by hand, or the two halves of the second are clapped together. These excitations produce a definite pitch corresponding to the fundamental mode of the closed-open resonator. If the other end of the single-piece resonator is slapped, the pitch differs by a musical third. If one half of the double-piece resonator is inverted, the pitch differs by one musical whole tone. If the other half is also inverted, the pitch differs by another whole tone.

Of the different ways to alter the pitch of a resonator, this is perhaps the most unusual. Note that the volume remains constant here; it is the *nonuniformity* that is effectively altered. In the lower-frequency case the resonator necks down toward the open end, and in the higher-frequency case it necks up. The different diameters were carefully calculated and machined to give musical intervals.[1] The effect can be qualitatively understood by noting that the neck-down configuration is similar to a Helmholtz resonator, which yields a low frequency in the limit of a narrow opening. The neck-up configuration effectively shortens the resonator in the limit of a narrow section near the closed end, which yields a higher frequency.

[1]Bruce Denardo and Steven Alkov, "Variably nonuniform acoustic resonators," *American Journal of Physics*, vol. 62, pp. 315-321 (1994). Bruce Denardo and Miguel Bernard, "Design and measurements of variably nonuniform acoustic resonators," *American Journal of Physics*, vol. 64, pp. 745-751 (1996).

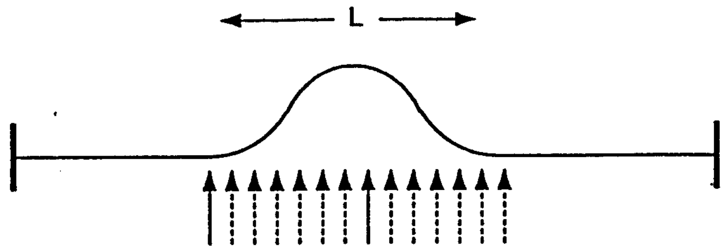
MW.5 Nonradiating wave source



Arrangement to demonstrate a nonradiating wave source. The current in the wire causes it both to glow and to experience a magnetic force due to the magnets. Motion of the wire visibly cools it. For nonradiating conditions, the motion is observed to be confined to the drive region.

A *nonradiating wave source* drives a medium over some region, but no radiation escapes the region due to complete destructive interference. Such sources have been predicted to occur in one and higher dimensions.[1] The one-dimensional case of a uniform force over some distance L of a string can be simply understood.[2] We consider the drive to be a uniform collection of a large number of point sources (see diagram below). If the frequency of the drive corresponds to wavelength L/2, L/4, L/6, ..., then the radiation cancels in pairs of the point sources, analogous to the minima of single slit diffraction. The diagram illustrates the fundamental nonradiating wave. There is a half-wavelength between the two point sources labeled by solid lines, and so their radiation cancels outside them. There is also a half-wavelength between all other similar pairs of sources, so no radiation escapes the driven region. The response inside the source has the appearance shown in the diagram. If the frequency is doubled, there is again a nonradiating state. The response inside has two humps now, with a node at the center.

A uniform force over length L with frequency corresponding to wavelength L/2 leads to complete destructive interference outside the driven region.



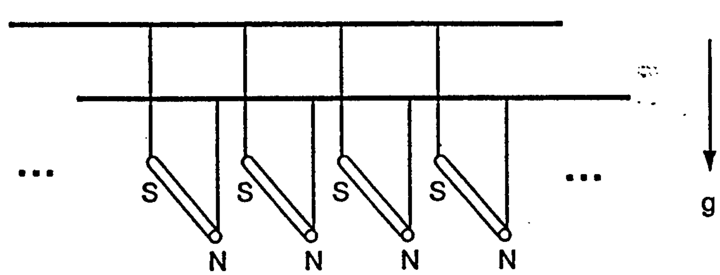
A nonradiating wave source can be demonstrated with waves on a wire that passes between the pole faces of an extended magnetic field (see above diagram). The Pasco magnets, which are 7.5 cm in width, can be juxtaposed if held by large clamps. An oscillatory current is passed through the wire, resulting in an oscillatory magnetic force perpendicular to the plane of the diagram. The force is approximately uniform between the pole faces and zero outside. Nichrome wire is used with sufficient current so that the wire glows when not in motion. The wire is dark in regions with sufficient motion, because this cools the wire. For the fundamental nonradiating state, the wire is dark over a region at the center of the magnets, and the wire glows uniformly outside the magnets. The second nonradiating state, which has two dark regions, is difficult to obtain.

[1]Michael Berry, John T. Foley, Greg Gbur, and Emil Wolf, "Nonpropagating string excitations," *American Journal of Physics*, vol. 66, pp. 121-123 (1998).

[2]Bruce Denardo, "A simple explanation of simple nonradiating sources in one dimension – Comment on 'Nonpropagating string excitations,' by M. Berry, J. T. Foley, G. Gbur, and E. Wolf," *American Journal of Physics*, vol. 66, pp. 1020-1021 (1998).

MW.6 Magnet pendulum lattice

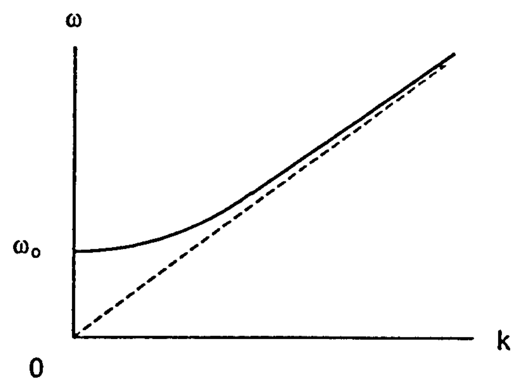
Section of a lattice of magnet pendulums, which exhibits longitudinal waves. (The vertical and other supports are not shown.)



Longitudinal waves can be demonstrated on a lattice of pendulums whose masses are magnets arranged as in the diagram above. The repulsive force between the magnets supplies the coupling. Note that if the pole orientations instead alternated, the lattice would be unstable for all but perhaps small-amplitude oscillations. It is interesting to exhibit waves to students and others without first revealing that the masses are magnets. Strange explanations for the coupling can ensue. In one case, it was argued that compressions and rarefactions of the air between the masses was responsible for the coupling!

If a wave packet is created by moving a pendulum on one end back and forth several times at a frequency roughly equal to the pendulum frequency, the packet is observed to propagate very slowly. This occurs because the waves are dispersive, and the phase velocity is large near the lowest frequency of propagation, while the group velocity is small. To understand this, we consider motion in the continuum limit, where the wavelength is large compared to the lattice spacing. For small-amplitude motion, the equation of motion for the displacement y(x,t) of the pendulum at position x and time t is $\partial^2 y/\partial t^2 - c^2\partial^2 y/\partial x^2 + \omega_o^2 y = 0$, where c^2 is a measure of the coupling strength and ω_o is the natural angular frequency of an uncoupled pendulum. The dispersion law for monofrequency waves, which is the relationship between the wavenumber k and angular frequency ω, is found by substituting $y = A\cos(kx - \omega t)$ into the equation of motion. The result is $\omega^2 = \omega_o^2 + c^2k^2$, which is shown in the graph below. Note that for infinite wavelength, where all of the pendulums are in phase with the same amplitude, the frequency is just the frequency ω_o of a single pendulum.

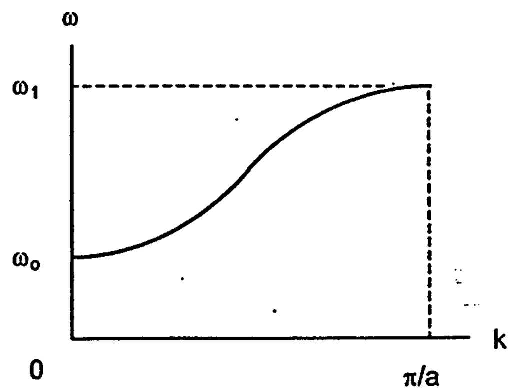
Dispersion law $\omega(k)$ for waves on a pendulum lattice in the continuum limit. The relationship asymptotically approaches $\omega = ck$.



In general, the phase velocity is $v_{ph} = \omega/k$ and the group velocity is $v_{gr} = d\omega/dk$. It is convenient to express the group velocity as $v_{gr} = (k/\omega)d(\omega^2)/dk^2 = (1/v_{ph})d(\omega^2)/dk^2$. Substituting the dispersion law $\omega^2 = \omega_o^2 + c^2k^2$ yields $v_{ph}v_{gr} = c^2$, which is well-known relationship that also occurs for waveguide modes. From the dispersion law, we find v_{ph} diverges in the limit $k \to 0$, so the group velocity must approach zero. Hence, a wave packet with frequency near ω_o moves very slowly.

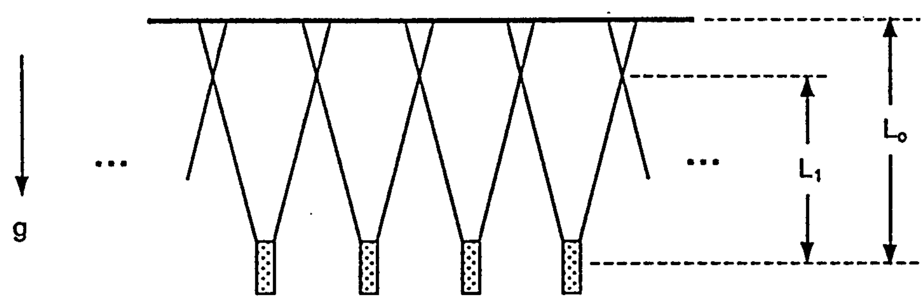
For the magnet pendulum lattice, it appears that wave packets move slowly even if the frequency is significantly greater than ω_o. This evidently occurs because the coupling is weak. For small values of c, the asymptotic line $\omega = ck$ law in the above graph rises slowly, so the group velocity $d\omega/dk$ remains small. The demonstration needs to be improved by adding a controllable drive.

The exact dispersion law for the magnet lattice can be calculated. Because there is both a lower cutoff mode (in which all the pendulums are in phase) and an upper cutoff mode (in which the pendulums are in antiphase), the dispersion law has the appearance shown in the diagram below. The wavelength of the lower cutoff mode is infinite (k = 0), and the wavelength of the upper cutoff mode is 2a ($k = \pi/a$), where a is the lattice spacing. The dispersion law is $\omega^2 = \omega_o^2 + (4c^2/a^2)\sin^2(ka/2)$, which is the standard mass-and-spring lattice dispersion law for ω^2 added to the constant ω_o^2. If the coupling c^2 is weak, then $\omega_1 \approx \omega_o$ and the group velocity $v_{gr} = d\omega/dk$ is small for all wavenumbers.



Exact dispersion law $\omega(k)$ for waves on a pendulum lattice. There are both lower and upper cutoff modes, so propagation occurs in a band of frequencies.

MW.7a Solitons on a pendulum lattice



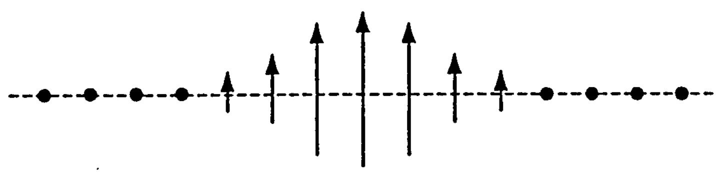
Section of a lattice of coupled pendulums whose motion is transverse. The support is oscillated vertically in order to parametrically excite the pendulums. Two types of standing (nonpropagating oscillatory) solitons are observed.

A lattice of 20 to 40 pendulums is constructed with string and machine nuts as shown in the above diagram. Suitable lengths are $L_o = 8.5$ cm and $L_1 = 6.0$ cm, with a 2.5 cm spacing between pendulums. A small knot is tied where the Vs overlap in order to ensure the coupling. The horizontal support is fixed by vertical rods that are attached to a shaker, which parametrically drives the lattice. The shaker can be a modified large loudspeaker. The lattice has a lower cutoff mode in which all of the pendulums move in phase with a common amplitude, and an upper cutoff mode in which the pendulums move in antiphase with a common amplitude (see MW.6). The linear frequencies of these modes are $f_o = (g/L_o)^{1/2}/2\pi$ and $f_1 = (g/L_1)^{1/2}/2\pi$, respectively. For the dimensions stated above, the parametric drive frequencies corresponding to the modes are $2f_o$ = 3.42 Hz and $2f_1$ = 4.07 Hz.

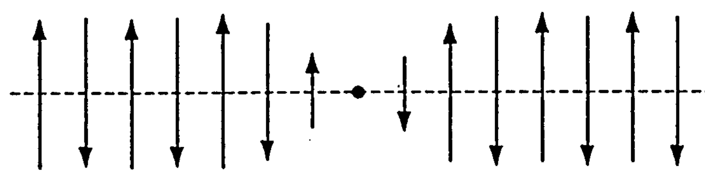
In the first demonstration, the drive is not employed. A meter stick is used to initiate the lower cutoff mode by releasing all of the pendulums from rest with a common amplitude. For a sufficiently large amplitude, the mode is observed to be unstable. The motion evolves into a number of localized oscillating "humps" that attenuate due to dissipation. The humps exhibit very little if any translational motion. These are examples of *solitons*, and are referred to specifically as *breathers*. Solitons are localized waves that act as particles, where nonlinearity is responsible for the localization.

A single breather (see diagram below) can be observed indefinitely by using the drive to overcome dissipation. The drive frequency should be slightly less than twice the linear frequency of the lower cutoff mode (3.38 Hz is suitable), and the peak-to-peak drive amplitude roughly 3/4 mm (350 mV pp into an APS shaker amplifier). A breather is initiated by hand, and will settle into the appropriate steady state profile. The breather can be moved by momentarily holding a pendulum in the tail, or by blowing air along the lattice.

Representation of the top view of a breather at a turning point of its motion. The pendulums outside the breather are at rest.



Representation of the top view of a kink at a turning point of its motion. The pendulums outside the kink have uniform amplitude.



Next, a different kind of standing soliton can occur in the upper cutoff mode. The drive frequency should be slightly less than the twice the linear frequency of the upper cutoff mode (4.05 Hz is suitable), and the peak-to-peak drive amplitude roughly 2-3 mm (1.0 V pp into an APS shaker amplifier). A duration of roughly 5-10 seconds is allowed to elapse so that the pendulums develop small-amplitude motion, and then the motion is amplified by reducing the drive frequency (reducing to 4.00 Hz is suitable). What is typically observed are regions of upper cutoff motion separated by a highly localized region of a pendulum at rest (see diagram immediately above). In some cases, the region may consist of two neighboring pendulums moving in phase with the same relatively small amplitude. These transitions between the upper cutoff regions are examples of "kink" solitons. As in the case of a breather, a kink can be moved by hand. If two kinks are brought near each other, they will annihilate.

The existence of the breather and kink states is remarkable because the system is responding in monofrequency modes that are not the usual uniform standing wave modes. To understand the localized modes, we first note that the drive serves to balance dissipation, so we ignore both of these. The key to understanding the breather is that it exists at a frequency *below* the linear frequency of the lower cutoff mode. We can consider the body of the breather as driving the tails, where linear theory applies. Because the frequency is below the linear cutoff value, the tails must exponentially evanesce with distance from the body, which is indeed observed. In the body, the nonlinearity of the pendulums causes the oscillations to soften (to be at a frequency less than the linear value) whereas the curvature of the profile produces an increase in frequency. The profile represents a competition between the nonlinearity and curvature such that there occurs a monofrequency response less than the linear frequency. It can be shown by weakly nonlinear theory that the profile satisfies a modified nonlinear Schrödinger equation that has hyperbolic secant solutions.[1]

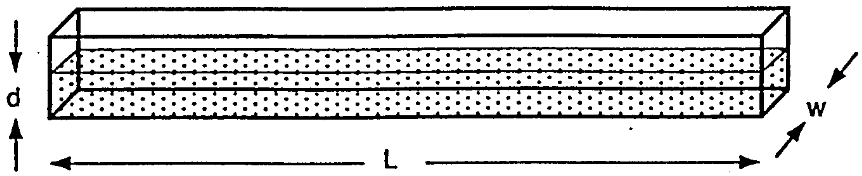
An understanding of the kink can be gained by considering the linear mode whose frequency is immediately below that of the upper cutoff mode. The profile of this mode is a sinusoid with a single node at the center of the lattice. Note that the effect of the curvature of the amplitude profile here is evidently to *lower* the frequency, which is the opposite of motion near the lower cutoff mode. (This occurs because the pendulums are in antiphase.) We now imagine that the amplitude of the mode is increased. The frequency will decrease due to the softening nonlinearity. The system maintains a spatially monofrequency response in two ways. Near the node, where linear theory applies, the curvature of the profile increases to reduce the frequency. In the wings (far from the node), where the nonlinearity operates, the curvature decreases. The ultimate result is that the profile becomes flat in the two wings, which oscillate 180° out-of-phase. The localized transition between these two regions is the kink, which behaves as a particle because it can be moved without altering the response in the wings. In particular, the kink is not influenced by the boundaries of the system. It can be shown by weakly nonlinear theory that the kink profile satisfies a modified nonlinear Schrödinger equation that has hyperbolic tangent solutions.[1,2] The pendulum lattice also reveals that kinks in modes other than the upper cutoff mode exist, and even that two different modes can coexist in juxtaposition.[2,3] The transition region between two different modes is referred to as *domain wall.*

[1]Bruce Denardo, *Observations of Nonpropagating Oscillatory Solitons* (Ph. D. dissertation, Department of Physics, University of California, Los Angeles, California, 1990).

[2]Bruce Denardo, Brian Galvin, Andrés Larraza, Alan Greenfield, Seth Putterman, and William Wright, "Observations of localized structures in nonlinear lattices: Domain walls and kinks," *Physical Review Letters*, vol. 68, pp. 1730-1733 (1992). Bruce Denardo and William B. Wright, "Structural properties of kinks and domain walls in nonlinear oscillatory lattices," *Physical Review E*, vol. 52, pp. 1094-1104 (1995).

[3]Bruce Denardo, Andrés Larraza, Seth Putterman, and Paul Roberts, "Nonlinear theory of localized standing waves," *Physical Review Letters*, vol. 69, pp. 597-600 (1992).

MW.7b Surface wave solitons



Channel of water that is vertically driven by a shaker (not shown). Steady-state surface wave breather solitons are observed for deep water, and kink solitons for shallow water.

The solitons observed in the parametrically driven pendulum lattice (MW.7a) are general states that can occur in other systems. Here we consider surface waves on a liquid. These waves have the linear (small-amplitude) dispersion law $\omega^2 = gk\tanh(kd)$, where ω is the angular frequency, g is the acceleration due to gravity, $k = 2\pi/\lambda$ is the wavenumber, and d is the depth. When kd is sufficiently large (deep liquid), the dispersion law can be approximated by $\omega^2 = gk$. When kd is substantially less than unity (shallow liquid), the dispersion law becomes dispersionless: $\omega = (gd)^{1/2}k$.

Consider a standing surface wave mode. Remarkably, the mode softens (natural frequency decreases with amplitude) for deep liquid and hardens (natural frequency increases with amplitude) for shallow liquid. The crossover occurs when kd = 1.058, so the crossover depth for the fundamental mode across a channel ($\lambda = 2w$ in the above diagram) is $d_c = 1.058w/\pi$.

The cross mode of the channel of liquid is similar to the lower-cutoff mode of the pendulum lattice. Because both modes soften, we thus expect breather solitons to occur in the channel of liquid when $d > d_c$. This is indeed the case.[1] As explained in the pendulum lattice demonstration (MW.7a), the monofrequency breather and kink solitons exist due to a balance between nonlinearity and curvature of the amplitude profile along the system. Kink solitons exist in the upper cutoff mode of the softening pendulum lattice. The curvature of the amplitude profile in this case has the opposite effect upon frequency compared to the normal case (for example, for a string or for the lower-cutoff pendulum lattice mode). Hence, the kink solitons that exist in the upper-cutoff mode of a softening system should exist in the lower-cutoff mode of a hardening system. Mathematically, this amounts to simply switching the signs of the nonlinearity and the effect of the curvature on frequency.[2,3] We thus expect kink solitons to occur in the channel of liquid when $d < d_c$, which is indeed the case.[2,3]

A surface wave channel can be constructed from clear acrylic. Suitable dimensions for the observation of a breather soliton are w = 3.0 cm, L = 38 cm, and d = 2.0 cm. So that a longitudinal mode is not excited by the parametric drive, the length L of the channel should be roughly midway between lengths corresponding to such resonances; that is, L/w should be roughly midway between integral values. Water can be used as the liquid. A small amount of a wetting agent such as Kodak Photo-Flo should be added to reduce the effect of pinning of the water along the walls. A small amount of flourescein or food coloring can be added so that the surface can be more easily seen. The parametric drive frequency should be roughly several percent less than twice the linear frequency of the cross mode, $2f_o = [gk\tanh(kd)]^{1/2}/\pi = 10.04$ Hz, where $k = \pi/w$. A suitable drive frequency is 9.85 Hz and peak-to-peak drive amplitude is roughly 1.0 mm (500 mV pp into an APS shaker amplifier). A small spatula can be used to initiate a breather, or the channel can be transversely tilted. As in the pendulum lattice, the uniform mode is unstable and will self-focus into one or more breather solitons.

The surface wave kink soliton is more difficult to obtain because of the requisite small depth ($d < d_c = 1.058w/\pi$), which causes substantial dissipation. Although the creation of a kink in the above channel is possible, it is difficult to accomplish and not easily seen. A better approach is to use ethyl alcohol in a large channel. Suitable dimensions are width w = 5.71 cm and length L = 76.2 cm. It is important that the width be uniform. Because alcohol slowly attacks acrylic, the alcohol should be removed from the channel after each use. A depth d = 1.0 cm can be used. It is very important to carefully level the channel, so that the depth is uniform. The drive frequency should be roughly a percent or two greater than twice the linear frequency of the cross mode, $2f_o$ = 5.23 Hz. A peak-to-peak drive amplitude of roughly 1.0 mm is appropriate. A kink can be initiated by imparting an angular impulse to the channel about a vertical axis not near the ends of the channel. A kink forms near the axis, and then slowly drifts due to nonuniformities. That a kink is a localized state can be dramatically demonstrated by moving a kink by inserting a small spatula parallel to the channel at the node of the kink, and then slowly moving the spatula. The kink can be removed by dragging it to an end of the channel. A pure cross mode then develops. It is also possible to create two kinks in the channel. By bringing them near each other, they do *not* annihilate as in the case of the pendulum lattice, but strongly repel. This is probably a Bernoulli effect due to the reduced velocity between the kinks.

[1]Junru Wu, Robert Keolian, and Isadore Rudnick, "Observation of a nonpropagating hydrodynamic soliton," *Physical Review Letters*, vol. 52, pp. 1421-1424 (1984). Junru Wu and Isadore Rudnick, "Amplitude-dependent properties of a hydrodynamic soliton," *Physical Review Letters*, vol. 55, pp. 204-207 (1985).

[2]Bruce Denardo, *Observations of Nonpropagating Oscillatory Solitons* (Ph. D. dissertation, Department of Physics, University of California, Los Angeles, California, 1990).

[3]Bruce Denardo, William Wright, Seth Putterman, and Andrés Larraza, "Observation of a kink soliton on the surface of a liquid," *Physical Review Letters*, vol. 64, pp. 1518-1521 (1990).

Sound

1. Standing waves: hot-wire tube

2. Propagation: Multipole source, unbaffled and baffled loudspeaker, and pulsed plane/waveguide in tube

3. Photoacoustics: strobe and glass with carbon

4. Thermoacoustics: Rijke tube and Hofler tube

5. Nonlinear acoustics in a propagating wave tube: distortion and shocks, single sine burst, sum and difference waves, and absorption by noise

6. Sound beam from a parametric array

7. Bernoulli attraction and Rayleigh disk

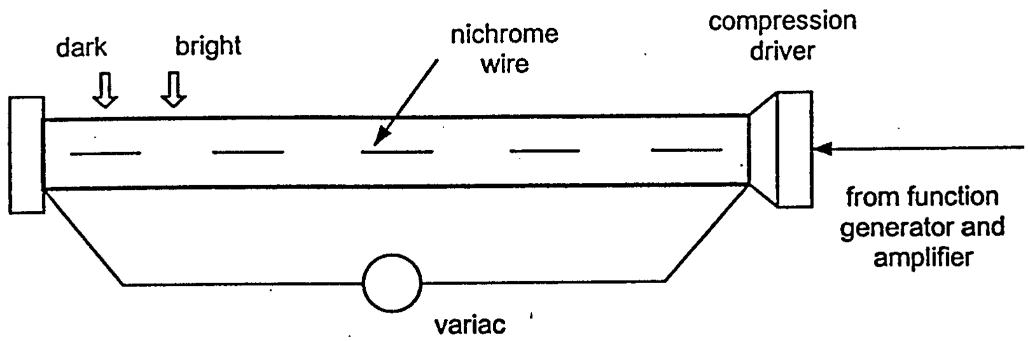
8. Kundt's tube: acoustic bunching and levitation

9. Acoustically driven jetting in Helmholtz resonators

10. Streaming: spinning cups

11. Radiation pressure: Crooke's and acoustic radiometers, acoustic Casimir effect, and flame tube

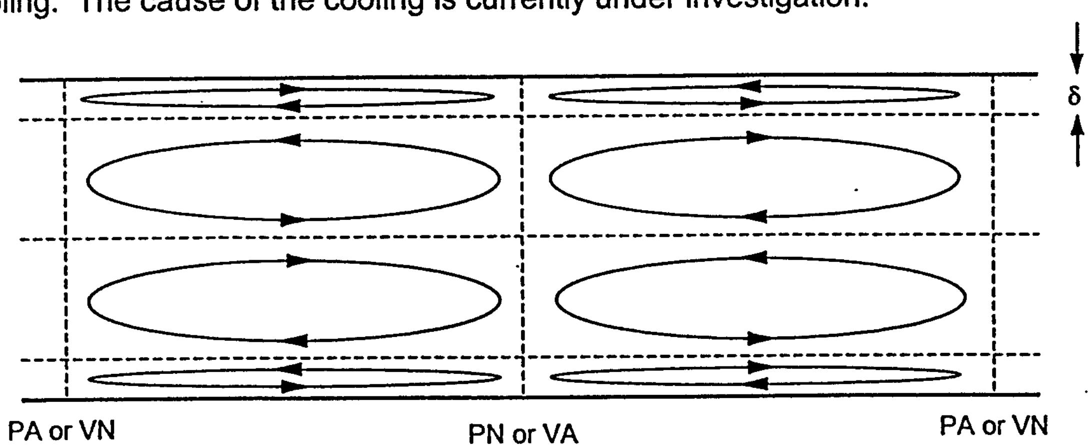
SO.1 Hot-wire standing wave tube



Standing wave tube with a hot wire on the axis. The wire uniformly glows when there is no sound. A high-amplitude standing wave causes alternate dark regions to occur.

A closed glass tube has a compression driver at one end. A nichrome wire lies along the axis of the tube. The voltage from a variac is adjusted so that the wire just visibly glows. A sufficiently high-amplitude standing wave cools the wire in regions of the velocity antinodes, causing the wire to be dark. A spatially-alternating bright and dark pattern thus occurs, allowing the standing wave to be seen.

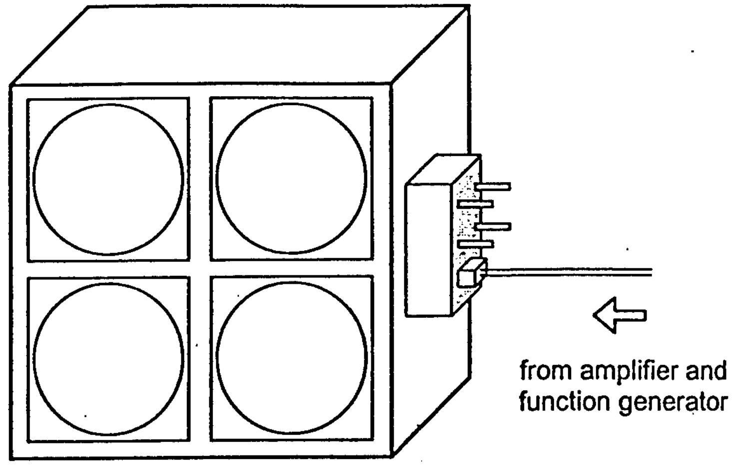
The mechanism of the cooling is not clear here, because the particle motion is *parallel* to the wire. Acoustic streaming (SO.10) may be the cause. Shown in the diagram below is the (steady) streaming motion for one-dimensional standing waves in a tube without a wire. The presence of the wire may cause an opposite flow within a viscous penetration depth of the wire, similar to the walls of the tube. This would cause cooler air to converge near a velocity antinode, and may thus be responsible for the cooling. The cause of the cooling is currently under investigation.



Steady streaming motion due to a standing wave in a tube. The viscous penetration depth is $\delta = (2\nu/\omega)^{1/2}$, where ν is the viscosity and ω is the angular acoustic frequency. PA and PN refer to pressure antinodes and nodes respectively, and VA and VN to velocity antinodes and nodes, respectively.

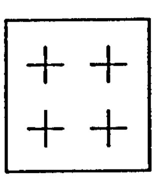
SO.2a Multipole source

Multipole source apparatus. Four identical loudspeakers are connected to a switching box, into which a signal is fed. The switches allow for each loudspeaker to be driven in phase or 180° out-of-phase relative to the signal. The system can thus be selected to act as a monopole, dipole, or quadrupole.

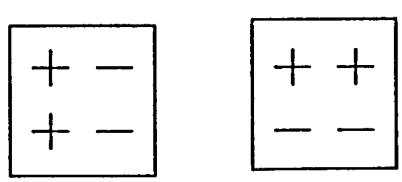


Four identical loudspeakers with enclosures are connected to a switching box, into which an audio signal is fed. The loudspeakers are connected in parallel with the signal, and the switches allow for the phase of the signal to each loudspeaker to be independently set in phase or 180° out-of-phase relative to the original signal. Each switch also has an "off" position. The frequency of the signal is chosen so that the wavelength is large compared to the apparatus. A frequency of 200 Hz (wavelength of 1.7 m) is suitable.

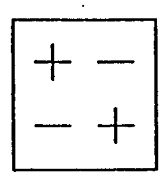
First, the polarities of the switches are selected to be the same. The four loudspeakers are thus all in phase and act as a monopole source (refer to the diagram below). Next, the polarities of two loudspeakers on a side are flipped, so that the loudspeakers now act as a dipole source. Because the wavelength is large compared to the spacing of the loudspeakers, the radiated sound in all directions is significantly reduced. Note that there are two possible arrangements for the dipole source (refer to the diagram below). Finally, the radiated sound is further reduced by adjusting the switches so that two dipoles are oppositely juxtaposed, so that the loudspeakers act as a quadrupole. The actions can be represented by holding up the palms of the hands next to each other, and moving them forward and backward. This is done first in phase to represent a monopole, and then out of phase (in two possible orientations) to represent a dipole. Finally, rotating the hands out of phase about the knuckles represents the quadrupole.



Monopole source

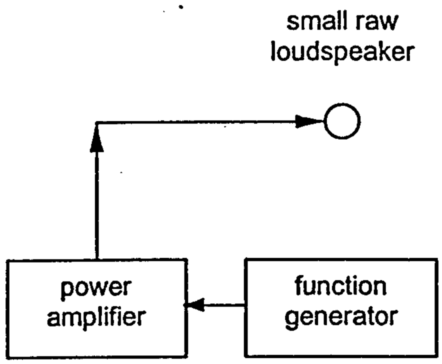


Dipole sources

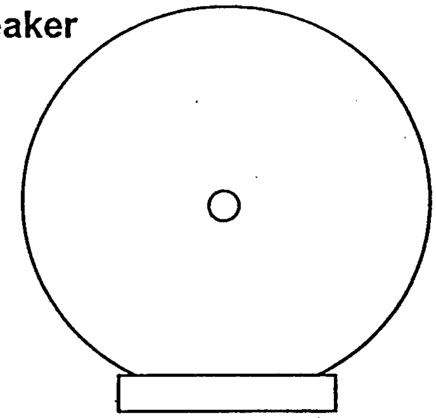


Quadrupole source

SO.2b Unbaffled and baffled loudspeaker



A small raw loudspeaker driven at a low frequency is inaudible or barely audible.



The loudspeaker is clearly audible when placed in the aperture of a large sheet of plywood which serves as a baffle.

A small (2 or 3-inch diameter) loudspeaker is driven with a low frequency (e.g., 200 Hz) such that the wavelength is large compared to the diameter of the loudspeaker. The loudspeaker is "raw;" that is, it is not part of an enclosure. The voltage to the loudspeaker is chosen such that very little, if any, sound is heard by the audience. With the loudspeaker remaining in operation, it is now pressed against the back of a large (3 or 4-foot diameter or side length) sheet of plywood, where there is a hole to accommodate the loudspeaker. The wood operates as a *baffle* for the loudspeaker, and the sound is now clearly audible.

The raw loudspeaker is a dipole source: When the cone creates a compression in front, it simultaneously creates a rarefaction in back. The wavelength being substantially larger than the size of the loudspeaker has two effects. First, there is substantial diffraction, so the sound produced on each side of the cone is propagated on the other side. The sound produced on each side thus interferes. Second, the large wavelength compared to the distance between the front and rear sources means that interference will tend to be destructive, so there is a small amplitude of sound. The effect of a large baffle is to change the dipole to a monopole, so that no interference occurs, which is responsible for the dramatic increase in amplitude.

The baffle causes the power of the sound to increase in both the front and rear. The source of this power must be the drive circuit. How does the drive circuit "know" to deliver more power? The answer is that the baffle alters the radiation impedance such that the drive circuit delivers more power. This can be quantitatively investigated with the apparatus. A power amplifier should be used to drive the loudspeaker so that the voltage is precisely the same for the two cases. (The voltage output of a function generator varies with the impedance.) A precision current meter then shows that the current surprisingly *decreases* when the baffle is employed. The effect is on the order of several percent. The only way that more power is delivered is if the phase between the current and voltage decreases a sufficient amount. That the phase indeed decreases can be demonstrated by using the signals to create a Lissajous pattern on an oscilloscope operating in the xy mode. This does not explicitly show that the power has increased, but makes it plausible. To complete the demonstration, a power meter should be used to verify that the power increases.

SO.2c Waveguide modes

Propagating wave tube. A closed PVC tube has two compression drivers attached to one end and a foam rubber absorber at the other end. The compression drivers are driven 180° out-of-phase with a several-cycle sine burst that produces a waveguide packet and a plane wave packet which travel at different speeds.

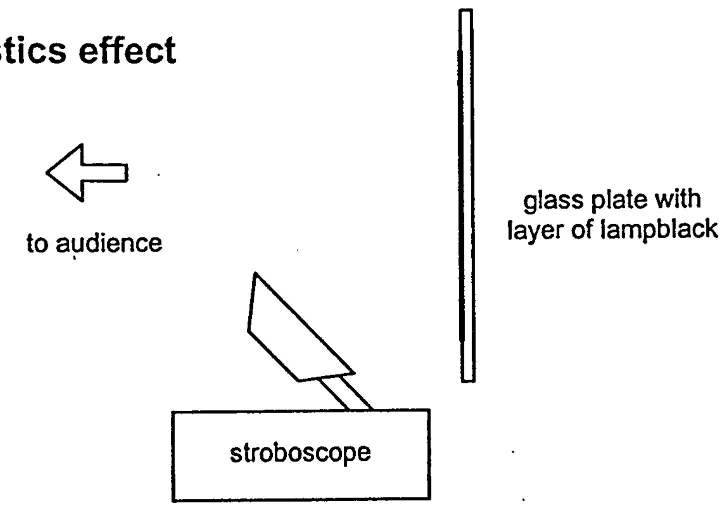
Plane waves of any frequency can normally occur in a waveguide, where the wave fronts are perpendicular to the waveguide. (Electromagnetic waves in a waveguide with a simply-connected cross section are an exception; plane waves cannot exist in this case.) In addition, modes where the wave fronts are not perpendicular to the waveguide can occur. These are composed of a standing wave perpendicular to the waveguide, and a propagating wave along the waveguide. Each such mode exists at frequencies above a *cutoff frequency*, which corresponds to the presence of only the standing wave component, and thus to the wave front being parallel to the waveguide. For a rigid acoustic waveguide with circular cross section of radius a, the cutoff frequencies are[1] $f_{mn} = cj'_{mn}/2\pi a$, where c is the speed of sound and j'_{mn} is the n^{th} zero of the Bessel function derivative $J_m'(r)$. The mode (m,n) has m interior pressure-node diameters and n pressure-node circles. The plane wave mode (0,0) can be included as the case $j'_{00} = 0$. The first several cutoff frequencies correspond to $j'_{10} = 1.841$, $j'_{20} = 3.054$, and $j'_{01} = 3.832$. As the frequency of a waveguide mode approaches the cutoff frequency from above, the phase velocity approaches infinity and the group velocity approaches zero. For frequencies much greater than the cutoff frequency, the phase and group velocities approach c. Hence, the group velocity of a waveguide mode is always less than the velocity c of a plane wave.

This difference is speeds can be demonstrated with the propagating wave tube shown above. (The same apparatus is used to demonstrate nonlinear acoustics effects in SO.5.) The inner diameter of the tube is approximately 2.0 inches, which yields the first cutoff frequency $f_{10} = 4.0$ kHz. The compression drivers are driven 180° out-of-phase by reversing the polarity of the wires to one of the drivers. A 5-cycle 4.5 kHz sine burst that repeats every half-second is input to the drivers. The microphone near the drivers shows a single pulse, but the microphone near the absorber shows that the pulse has split into two, corresponding to the plane wave and the (1,0) wave. The plane wave is excited due to asymmetries in the drivers and the driver geometry. If the frequency is increased to 5.5 kHz, the time interval between the pulses is observed to decrease.

[1]Philip M. Morse, *Vibration and Sound* (Acoustical Society of America, 1981), pp. 397-399. The zeros of the Bessel function derivatives are $j'_{mn} = \pi\alpha_{mn}$.

SO.3 Photoacoustics effect

The clicking of a stroboscope is substantially louder when the light strikes a plate upon which carbon has been deposited.



A glass plate with a layer of "lampblack" (particles containing carbon) is required here. This can be obtained by carefully holding a lit candle underneath the plate so that the flame touches the plate, and slowly and patiently moving the flame. The carbon does not adhere well to the glass, so it should not touched by hand or hit with an object. To help prevent this, and to signify which side is carbonized, it is useful to write a statement such as "Carbon Side – Do Not Touch" on the side.

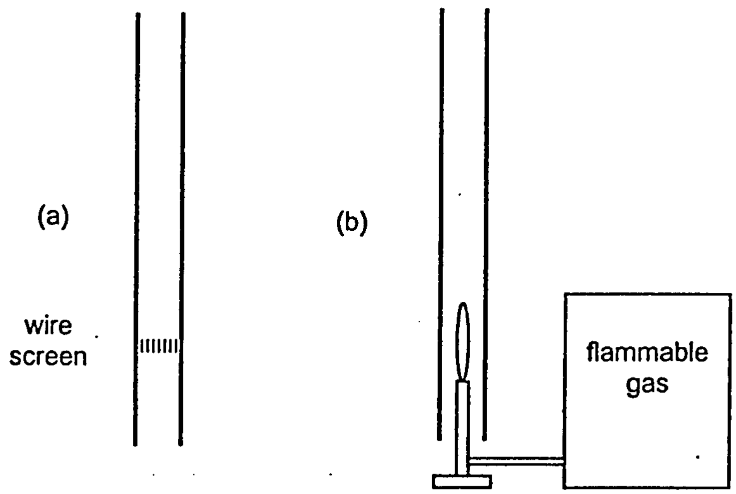
In the demonstration, a stroboscope is operated at a frequency of roughly 10 Hz. First, the stroboscope is operated normally and the audience is asked to note the loudness of the sound produced. A glass plate without carbon is then placed as shown in the diagram, and the audience is asked to observe that the sound altered very little if at all. Next, the glass plate with lampblack is brought near the stroboscope, and the loudness substantially increases. This occurs because the carbon absorbs the light, causing an increase in temperature. The resultant abrupt thermal expansion produces the sound. The field of *photoacoustics* deals with such phenomena.

It is interesting to reverse the orientation of the plate, so that the carbon is on the opposite side. A comparison is performed with the plate without carbon, and the sound to the audience is found to be only slightly louder. Next, all of the apparatus are turned so that the carbon side now faces the audience, and the comparison is performed again. The sound is observed to be louder, although it is not as loud as in the initial demonstration above. Because the glass is much more dense than the air, the carbon expands in the direction of the air, and so the sound is emitted in that direction. The reduction of this sound when the plate reversed is not due to the reflection of the light from the glass, because this is small. The reduction probably occurs because the infrared part of the light spectrum is primarily responsible for the heating of the carbon, and infrared radiation is strongly absorbed by normal glass. The thickness of the carbon layer may also play a role in the reduction of the sound.

Pasco carbon paper for use in electric field mapping experiments, as well as carbon paper for use with typewriters, do *not* yield sound. This may be due to a lack of sufficient carbon or to damping caused by the paper.

SO.4a Rijke tube

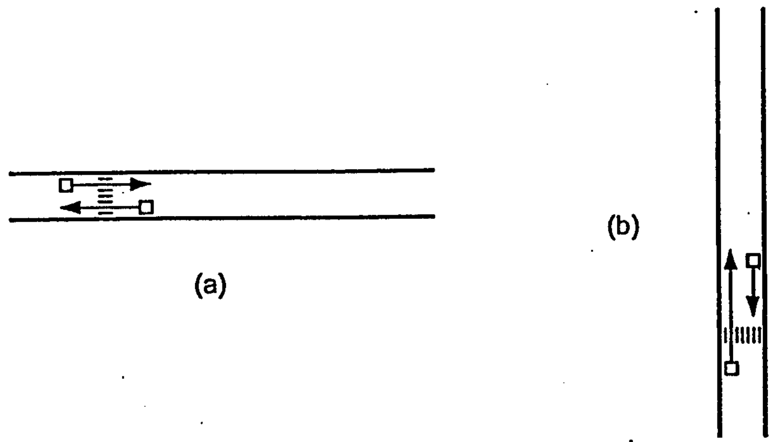
Heat-driven standing wave of sound in a open-ended tube: (a) heat supplied by a wire screen that was heated by a Bunsen burner, and (b) heat directly supplied by a Bunsen burner.



This effect was discovered in 1859 by Rijke.[1,2] A vertical aluminum tube open at both ends has several layers of wire screen located roughly one-quarter tube length from the bottom. The wire is fully heated with a Bunsen burner, which is then removed. A very loud sound of definite pitch (the fundamental of the tube) is subsequently heard until it eventually ceases as the wire cools. Several interesting demonstrations can be done while the wire is sufficiently hot. The sound ceases when either end of the tube is covered, and recurs when uncovered. The sound also ceases when the tube is rotated to a horizontal orientation, and recurs when the tube is returned to the original vertical orientation. Sound can be produced in the horizontal orientation if the tube is translated along its length if the screen is in the forward direction. (This can be achieved by the demonstrator running with the tube.) These demonstrations show that convection is necessary for the sound to be produced. The sound also ceases if the tube is inverted, so that the wire is in the top half. For the normal orientation of the tube, the sound can alternatively be produced without a wire screen by directly heating the air with the Bunsen burner.

The effect can be understood as follows. We imagine that the fundamental acoustic mode of the tube is excited. The problem is to explain how heat from the wire screen is fed into this mode to maintain it in the presence of dissipation. There is evidently more heat added to the air during a compression rather than an expansion. Consider first a horizontal tube [refer to diagram (a) below]. Due to the acoustic oscillations, a small parcel of gas is compressed and heated as it translates toward the middle of the tube, and expanded and cooled as it translates away from the middle. A parcel that passes through the screen is heated equally whether the motion occurs during a compression or expansion. Hence, the heat cannot maintain the oscillations.

Convection breaks the symmetry that occurs for the horizontal tube. The motion of a parcel is now the superposition of oscillatory and steady upward motions. Some parcels are thus only heated during compression [refer to diagram (b) above], and the oscillations can be maintained.



In (a), a small parcel of gas that passes through the wire screen is heated equally whether the motion is during a compression or expansion of the acoustic wave. In (b), convection causes some parcels to be heated only during a compression.

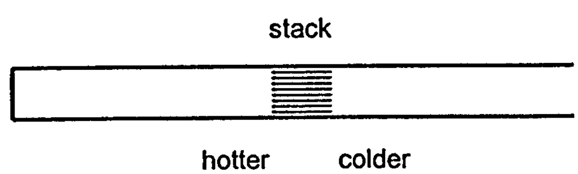
The explanation reveals why sound does not occur when the tube is inverted, because in this case heat is added during the expansion rather than the compression. Sound can be produced with the wire screen in the upper half of the tube if the wire sufficiently *cools* the air, and if there is an upward flow of hot air.[2] It may be possible to achieve this by pouring liquid nitrogen onto the wire while lightly heating the air at the bottom of the tube.

[1]A. B. Pippard, *The Physics of Vibration* (Cambridge, New York, 1989), pp. 343-345.

[2]John William Strutt (Lord Rayleigh), "The Theory of Sound," vol. II (Dover, New York, 1945), pp. 231-234.

SO.4b Hofler tube

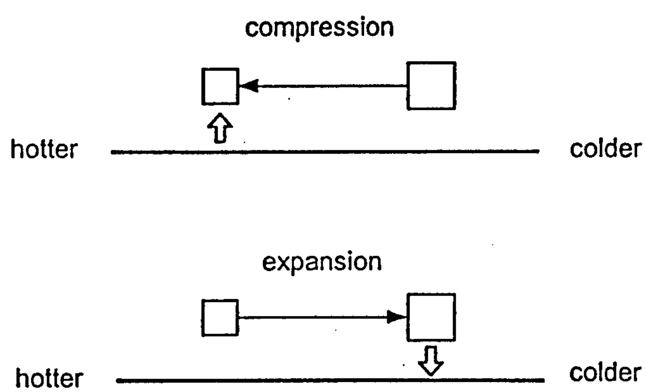
Standing wave of sound driven by a gradient in temperature in a closed-open tube. Air can oscillate longitudinally in the stack, which has an externally maintained temperature gradient along it.



This thermoacoustic device was invented by Thomas Hofler (Physics Department, Naval Postgraduate School).[1,2] The essential geometry is shown in the diagram above. Roughly midway in a closed-open metal tube is a stack composed of parallel strips of material that has high heat capacity but low thermal conductivity. Stainless steel can be used. Each end of the stack is maintained at a different temperature, where the end towards the closed end of the tube is at a hotter temperature than the other end. The hotter side can be heated with a Bunsen burner while the colder side is at room temperature, or the colder side can be cooled with liquid nitrogen while the hotter side is at room temperature. A very loud sound whose pitch corresponds to the fundamental (quarter wavelength) mode of the tube is heard.

To understand the effect, suppose that the fundamental acoustic mode is excited, and consider a small parcel of air that is oscillating in the stack (see diagram below). As the parcel is compressed, its temperature rises. If the temperature of the stack at the location of the parcel rises with distance more rapidly, heat will flow from the stack to the parcel, thus driving the acoustic mode. Similarly, heat is removed when the parcel is expanded and cooled, which also drives the mode. The acoustic oscillations can thus be maintained if the temperature gradient of the stack is greater than a threshold value.

Small parcel of air oscillating in the stack, one plate of which is shown. As the parcel is compressed, heat is added from the plate. As the parcel is expanded, heat is absorbed by the plate.

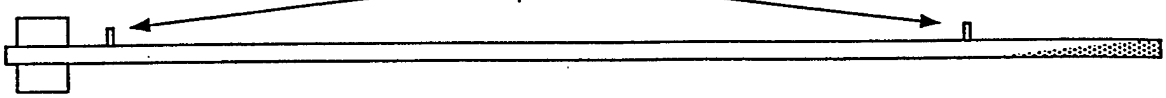


The thermoacoustic mechanism here is *reversible*. If the fundamental mode is driven (for example, by a loudspeaker at the closed end), a steady thermal gradient will occur along the stack. This is the basis of thermoacoustic refrigeration.

[1]J. C. Wheatley, T. Hofler, G. W. Swift, and A. Migliori, "Understanding some simple phenomena in thermoacoustics with applications to acoustical heat engines," *American Journal of Physics*, vol. 53, pp. 147-xxx (1985).

[2]J. C. Wheatley and A. Cox, "Natural engines," *Physics Today*, vol. 38, pp. 50-xxx (August, 1985).

SO.5 Nonlinear acoustics in a propagating wave tube

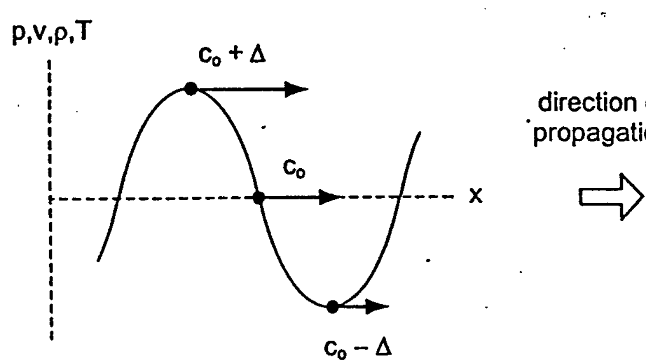


Closed PVC tube has two compression drivers attached to one end and foam rubber absorber at the other end. The finite-amplitude effects of distortion, shocks, sum and difference frequency production, suppression of sound by sound, and absorption of sound by noise are demonstrated for pure tones.

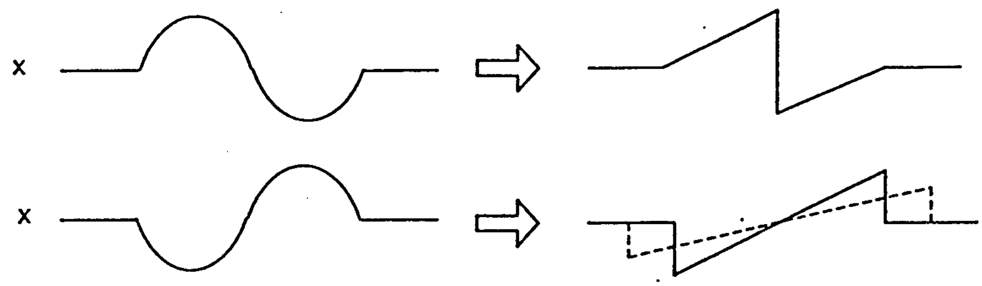
Several sections of 2-inch inner-diameter schedule-40 PVC pipe are joined together to make a pipe roughly 40 feet long. Two high-power compression drivers are attached at one end, and are driven by function generators and a dual-channel power amplifier. At the other end is several feet of a foam rubber absorber. The tube is closed at both ends. Microphones are placed as shown in the diagram, and are fed to preamplifiers and then an oscilloscope. In all of the demonstrations, the drivers are at frequencies below the first cutoff frequency (4.0 kHz), so that only plane waves propagate.

In the first demonstration, one or both of the drivers is made to produce a low-amplitude pure tone. The waveforms on the oscilloscope are sinusoidal, and some attenuation is observed at the second microphone if the sensitivities of the microphones and preamplifier gains are matched. As the drive amplitude is increased, the output of the first microphone remains sinusoidal, but that of the second microphone exhibits distortion in the form of a steepening of the regions of positive slope. The presence of higher harmonics can be seen on a spectrum analyzer. At a sufficiently high amplitude, the steeper slope at the pressure zero crossing becomes infinite. At a slightly greater amplitude, a shock front is observed. Ringing of the microphone can also be observed, and can be reduced by filtering. The distortion and eventual shocking are due to compressions traveling faster than expansions (see diagram below). This occurs for two reasons: the speed of sound is greater for higher temperature, and the speed of sound is boosted by the particle velocity.[1]

Compressions travel faster than the linear speed of sound c_o, and expansions travel slower. A sinusoidal wave thus distorts, eventually becoming a sawtooth wave with shock fronts.



In the second demonstration, the input is a single sine burst that repeats at a low rate. At high amplitudes, the burst evolves into either a single shock front or an N wave, depending upon the polarity of the burst (refer to the diagram below). The N wave lengthens in time (dashed lines in the diagram) because the negative-pressure shock front moves more slowly than the positive-pressure shock front.



Depending upon its polarity, a single high-amplitude sine burst evolves either into (a) a single shock front, or (b) an N wave. Shown here are the variations in space (rather than time), which can be displayed on an oscilloscope by using the input invert feature.

In the third demonstration, the drivers are driven at different frequencies (750 Hz and 2.0 kHz are suitable). At small amplitudes, the spectrum analyzer shows that the response at the first and second microphones is simply the superposition of the two tones. At greater amplitudes, however, many more peaks occur at the second microphone. The larger of these can be shown to be the sum and difference frequencies of the primary waves. The other peaks involve sum and difference frequencies of various combinations of the primary and secondary waves.

Suppression of sound by sound is demonstrated next. One driver emits a high-amplitude tone ("pump") at low-frequency (750 Hz is suitable), and the other a low-amplitude tone ("signal") at a high frequency just below cutoff (4.0 kHz). Due to the creation of sum and difference waves, energy is removed from the signal as it propagates. Remarkably, the amplitude varies in space as the absolute value of the $J_o(x)$ Bessel function, where the distance to the zeros is inversely proportional to the amplitude of the pump.[1] Just after the first zero, the amplitude of the signal increases with x due to a restitution of energy from the sum and difference waves. The zero can be exhibited by observing the sound at the second microphone with a spectrum analyzer. As the amplitude of the pump is slowly increased, the amplitude of the signal falls to zero and then rises.

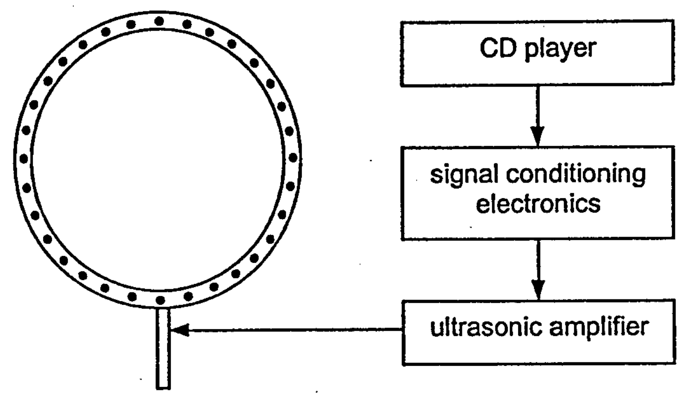
In the final demonstration, one of the drivers is driven with noise in a band of 0.5-2.0 kHz, while the other is driven with a pure tone at a higher frequency. It is convenient to use a brickwall filter to achieve an abrupt band of noise. To maximize the effect, the pure tone should be just below the cutoff frequency and as great an amplitude as possible without significant distortion. The output of the second microphone is fed to a spectrum analyzer. As the noise intensity is increased, the amplitude of the pure tone is observed to decrease, showing that finite-amplitude noise absorbs energy of a wave. This is due to the sum and difference waves produced by the wave with all of the components of the noise. Remarkably, this causes the wave to attenuate with distance not as an exponential but as a *gaussian*.[2]

[1]Mark Hamilton, "Fundamentals and Applications of Nonlinear Acoustics," in *Nonlinear Wave Propagation in Mechanics*, AMD-Vol. 77, edited by T. W. Wright (American Society of Mechanical Engineers, New York, 1986), pp. 3-6, 8-9. Mark F. Hamilton and David T. Blackstock, editors, *Nonlinear Acoustics* (Academic, San Diego, 1998), pp. 66-79, 82-84.

[2]Andrés Larraza, Bruce Denardo, and Anthony Atchley, "Absorption of sound by noise in one dimension," *Journal of the Acoustical Society of America*, vol. 100, pp. 3554-3560 (1996).

SO.6 Sound beam from a parametric array

A ring of transducers transmits high-intensity ultrasound. Due to the interaction of sound with sound, a narrow beam of audible sound is produced.



This remarkable device was developed by a company.[1] A 20-cm diameter ring has many small ultrasonic transducers mounted on it. The transducers, which operate at roughly 30 kHz, are connected in parallel to two wires from an amplifier, signal conditioner, and CD player. An audio microphone and preamplifier, or some other source, can be substituted for the CD player. As a result of nonlinear interactions of the sound in the air, the original audio signal is broadcast in a beam with distinct boundaries and no side lobes. The beam can be reflected from walls.

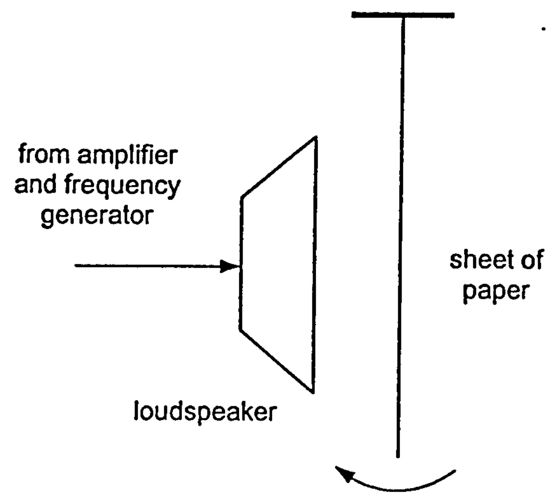
To understand the effect, consider a high-amplitude wave consisting of the relatively slow amplitude modulation of a monofrequency wave (carrier). Nonlinear interactions occur over roughly the attenuation length of the carrier, and can be shown to produce a wave whose pressure is proportional to the second derivative of the square of the modulation function.[2] This is referred to as *self-demodulation*. The carrier attenuates relatively rapidly, leaving only the demodulated wave. This wave is driven over a length that is substantially greater than the size of the primary source. The resultant beam is highly directional without side lobes. This is the principle of the *parametric array*, which is normally realized by two primary waves of slightly different high frequencies.[2] Nonlinear interactions drive a wave whose frequency is the difference of the two primary frequencies. The effect is as if there were an in-line array of linear sources whose phase variation and amplitude taper are precisely that which yield a highly directional on-axis beam with no side lobes. This occurs automatically in the case of nonlinear interactions of two collinear primary waves or a single amplitude-modulated primary wave. In addition to adding the high-frequency carrier, the signal conditioner of the above device evidently "predistorts" the signal so that the self-demodulated sound is nearly identical to the original audio signal.

[1]American Technology Corporation (San Diego, California). Web site: http://www.atcsd.com.

[2]Mark Hamilton, "Fundamentals and Applications of Nonlinear Acoustics," in *Nonlinear Wave Propagation in Mechanics*, AMD-Vol. 77, edited by T. W. Wright (American Society of Mechanical Engineers, New York, 1986), pp. 13-16. Mark F. Hamilton and David T. Blackstock, editors, *Nonlinear Acoustics* (Academic, San Diego, 1998), pp. 246-252.

SO.7a Bernoulli attraction

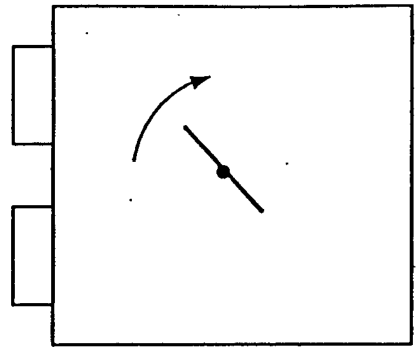
Side view of a sheet of paper supported in front of a loudspeaker. When the loudspeaker is driven so that a low-frequency high-amplitude sound wave is emitted, the paper is attracted to the loudspeaker.



A suspended sheet of paper is attracted to a loudspeaker that is driven with sufficient amplitude at a low frequency (200 Hz is suitable). Because the flow velocity is greater on the loudspeaker side of the sheet, Bernoulli's law implies that the steady pressure is reduced there. This occurs regardless of the direction of the flow, so the sheet is attracted to the loudspeaker.

SO.7b Rayleigh disk

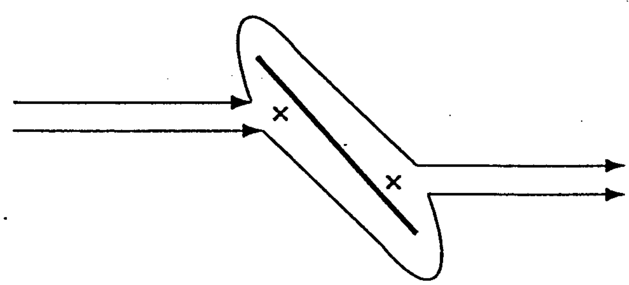
Top view of a note card suspended by string in an acoustic chamber. When the loudspeakers (rectangles) drive a high-amplitude standing wave mode that has a velocity antinode at the center, the card rotates to become perpendicular to the flow.



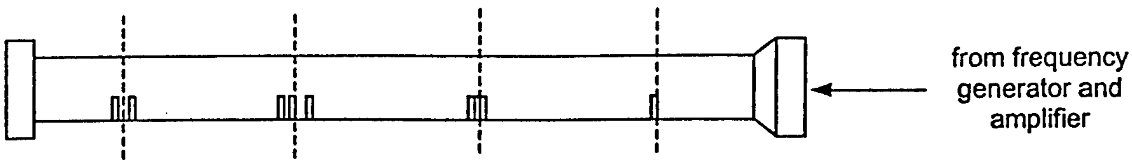
Illustrated above is a Bernoulli effect in addition to the attraction of a sheet of paper to a loudspeaker (SO.7a). A note card is suspended from a string that is taped to the center of the top of an acoustic chamber. In equilibrium with no sound, the note card is at an angle as shown in the diagram above. A standing wave mode is driven in one direction of the chamber such that a velocity antinode occurs at the center. This corresponds to modes with an *odd* number of half-wavelengths across the chamber, which includes the fundamental mode. For sufficiently high amplitudes of the mode, the note card rotates such that it is perpendicular to the particle displacement. The note card is referred to as a *Rayleigh disk* in this case.

To understand the torque on the note card, consider the flow (see diagram below). The separation points of the flow occur must occur above and below the center of the card, as shown. By Bernoulli's law, the lack of velocity at these points implies a greater steady pressure. The same occurs when the flow reverses. Hence, there is a steady torque that causes the disk to orient perpendicular to the flow.

Flow around a Rayleigh disk. By Bernoulli's law, the pressure is greater at the stagnation points (x's). There is thus a clockwise torque on the disk.



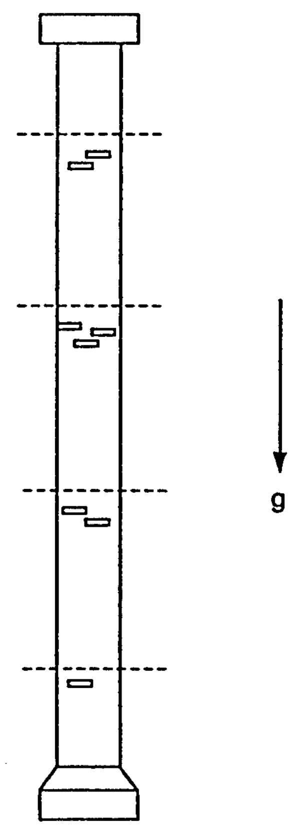
SO.8 Kundt's tube: Bunching and levitation



Kundt's tube, consisting of a closed clear acrylic tube with a compression driver at one end. When a high-amplitude standing sound wave is driven, small styrofoam disks tend to bunch at velocity antinodes (dashed lines).

The Kundt's tube is a dramatic Bernoulli demonstration in addition to Bernoulli attraction (SO.7a) and the Rayleigh disk (SO.7b). High-amplitude standing waves are driven by a compression driver at one end of a closed clear acrylic tube. Small styrofoam disks tend to bunch at velocity nodes (pressure antinodes) because the time-averaged pressure is least there. On either side of a velocity antinode, there is a time-averaged pressure gradient that forces the disks to the velocity antinode. In addition, the Rayleigh disk effect causes the disks to orient perpendicular to the oscillations.

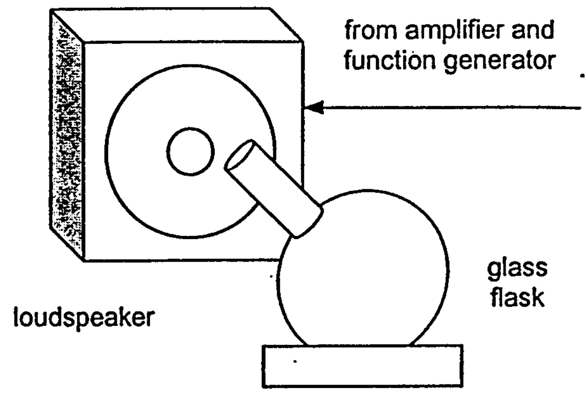
If the tube is then tilted so that it is vertical, the disks can be levitated. When this occurs, the upward force due to the time-averaged pressure gradient balances the downward gravitational force. The disks thus levitate below the velocity antinodes (refer to the dashed lines in the diagram to the right).



If the apparatus is tilted vertically, the disks can be levitated.

SO.9 Acoustically-driven jetting

A glass flask is driven at its Helmholtz resonance frequency with a loudspeaker. At sufficiently large amplitudes, a lit match placed in front of the neck will be extinguished.

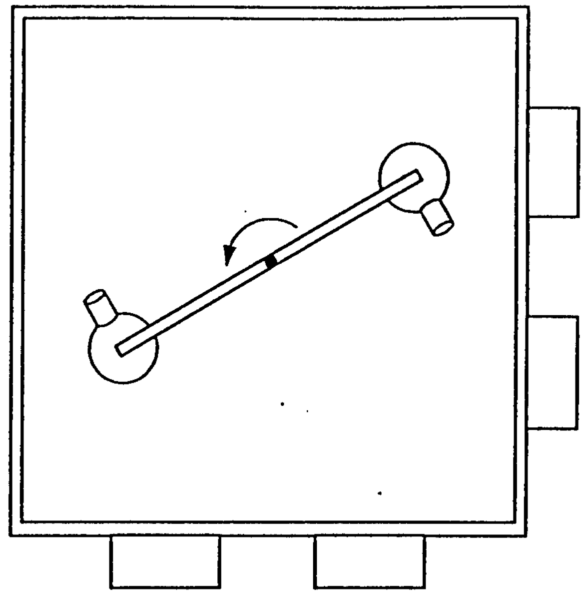


An open glass flask serves as a Helmholtz resonator in this demonstration. First, a small microphone connected to an oscilloscope is placed inside the flask, and the frequency of the loudspeaker is varied until the resonance value is determined. The amplitude should not be large, so that distortion of the microphone is prevented, and so that the sound is not unpleasant to the demonstrator and audience. For a typical one-liter flask, the resonance frequency is approximately 180 Hz. Next, the microphone is removed, and the amplitude is increased to a high level. A match is lit and brought near the front of the neck of the flask. A steady flow of air jetting from the resonator can extinguish the flame. To show that the sound oscillations cannot extinguish the match, the resonator can be removed and a lit match brought near the loudspeaker.

The steady flow can be explained as follows. By Bernoulli's law, the large-amplitude Helmholtz oscillations of the air in the neck lower the steady pressure there. This creates a steady in-flow of air, which must be accompanied by an out-flow. For a pipe immersed in a fluid, when fluid is sucked into a pipe, it tends to enter from all directions. However, when it is blown out, it tends to form a jet perpendicular to the plane of the end of the pipe. The flow is symmetric only for very small velocities. The out-flow jetting from the Helmholtz resonator can be observed just prior to the match being extinguished, or the drive amplitude can be reduced so that the match remains lit but the flame is pushed outward.

The jetting effect can be utilized to create an acoustic motor. Two small Christmas ornaments with their stems removed are mounted on either end of an arm that is pivoted at its center (refer to the diagram below). The ornaments serve as Helmholtz resonators. The assembly is placed in a closed clear acrylic box with loudspeakers attached to the sides. The top plate is hinged so that it can be open and closed. The horizontal cross section of the box is a square with height roughly one-third the side length of the square. To help achieve large acoustic amplitudes, the box is constructed such that its fundamental acoustic mode has approximately the same frequency as the Helmholtz resonance of the ornaments. The inside length of a side is $a = 16$ inches, so that the fundamental acoustic mode [that is, the (1,0) or the (0,1) mode] has approximate frequency $f = c/2a = 420$ Hz.

Top view of rotating Christmas ornaments in a closed clear acrylic box with loudspeakers attached to the sides. The ornaments rotate counterclockwise due to acoustically-driven jetting.

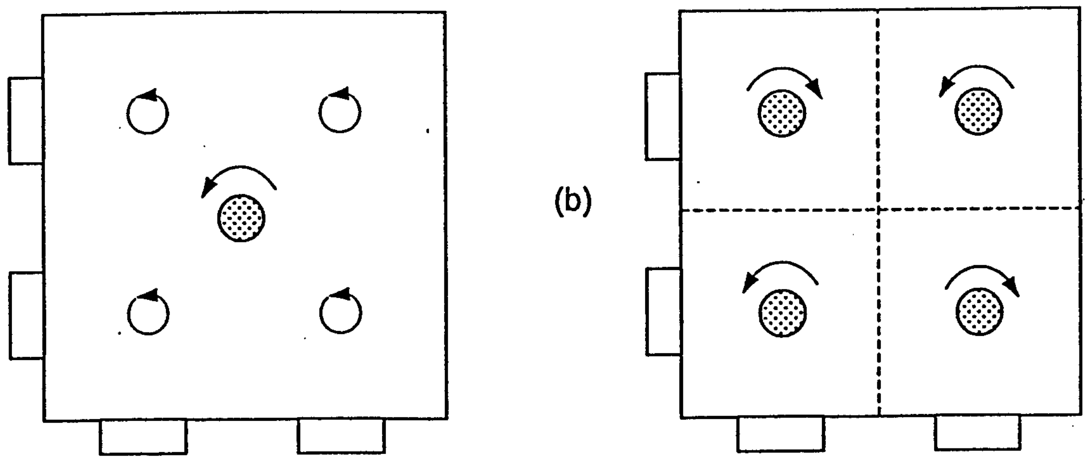


With the ornaments assembly initially at rest, the loudspeakers are driven at resonance and the assembly accelerates to a fast angular velocity. The demonstration can be repeated with the assembly being given an initial angular velocity in the opposite direction of the torque due to the jetting. Before the loudspeakers are turned on, the assembly moves with nearly constant angular velocity. After the loudspeakers are turned on, the assembly decelerates and reaches its terminal angular velocity as before.

During operation, the sound level is heard to be modulated by the rotation of the ornaments. This may occur because the resonance frequency of the cavity is modified depending upon the orientation of the ornaments. Because the drive frequency is fixed near this modulated resonance frequency, the response amplitude will increase when the resonance frequency approaches the drive frequency, and decrease when it recedes from it.

SO.10 Spinning cups

(a) (b)

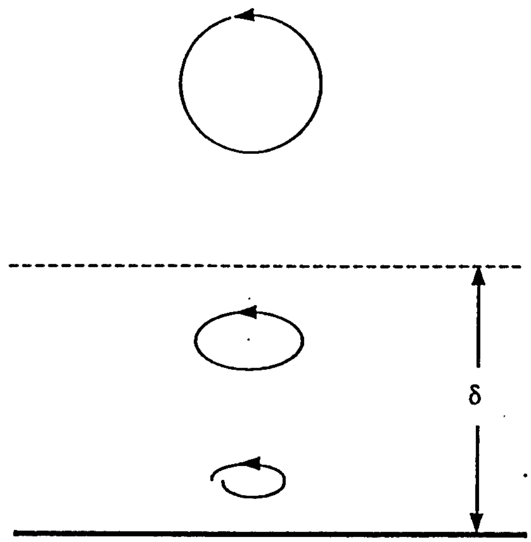


Top view of a closed acoustic chamber with styrofoam cups (shaded disks) supported on spindles. The cups spin when appropriate modes are driven at high amplitude by pairs of loudspeakers (rectangles) that are 90° out-of-phase.

A closed clear acrylic chamber has pairs of drivers on adjacent sides, as shown above. In (a), a styrofoam cup pivoted on a spindle is at the center. The pairs of loudspeakers are made to drive the fundamental mode in each direction with the same amplitude but 90° out-of-phase. The particle motion is thus uniformly circular. The velocity (circular arrows in the diagram) is $\mathbf{v} = v_o[\cos(\pi x/L)\cos(2\pi ft)\mathbf{i} + \cos(\pi y/L)\sin(2\pi ft)\mathbf{j}]$, where the origin is at the center, L is the length of each side, $f = c/2L$, and the x displacement is assumed to lead the y displacement. One might think that this would cause the cup to spin in the opposite direction due to viscous drag. However, the cup spins in the *same* direction has the particle motion. In diagram (b), four cups are employed and the drive frequency f is doubled. The particle velocity now alternates between quadrants, and the cups again spin in the same direction as in each quadrant.

The effect is due to *acoustic streaming*, which is a large-scale steady flow induced by acoustic motion. Streaming can occur when there is attenuation due either to bulk motion or to boundaries, so that the energy of sound wave is attenuated but the momentum is not. The streaming that causes the cups to spin is unusual in that it arises from circular rather than oscillatory acoustic motion. The following is a paraphrase of a physical explanation in Ref. 1. Outside of the viscous layer of the boundary of the cup, the particle motion is circular. Inside, however, the loop becomes smaller due to viscosity and is flattened due to the presence of the boundary (see diagram below). If the loop is deep within the layer, the half of the motion closer to the boundary is subjected to more viscous damping than the other half. This reduces the size of the former, which causes the loop not to be closed and results in a shift of the particle motion (to the left in the diagram). The accumulation of these shifts over many cycles constitutes a drift motion or streaming, which exerts a torque on the cup.

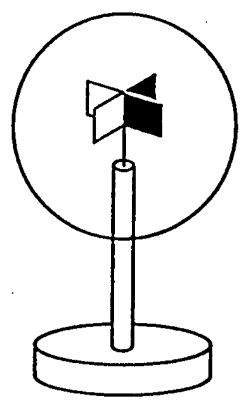
Particle motion outside and inside the viscous layer of a boundary, from Ref. 1. The viscous penetration depth is $\delta = (\nu/\pi f)^{1/2}$, where ν is the viscosity. Near the boundary, the loops are not closed, which results in a drift motion to the left.



[1]Chun P. Lee and Taylor G. Wang, "Near-boundary streaming around a small sphere due to two orthogonal standing waves," *Journal of the Acoustical Society of America*, vol. 85, pp. 1081-1088 (1989).

SO.11a Crooke's radiometer

Crooke's electromagnetic radiometer. When exposed to light of sufficient intensity, the vanes spin with the white side leading the black (counterclockwise viewed from above).

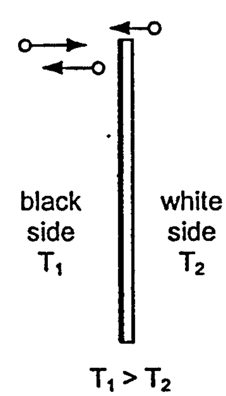


Crooke's electromagnetic radiometer is a well-known physics demonstration and toy. A glass bulb contains a freely rotating assembly of four vanes, where one side of each vane is black and the other side is white (refer to the diagram above). When exposed to light of sufficient intensity, the vanes rotate. Light has momentum, and thus exerts *radiation pressure* when it strikes a surface. Because light is absorbed by the black sides and reflected by the white sides, there is a greater momentum transfer (nearly twice) on the white sides. Due to this imbalance in radiation pressure, the vanes should rotate from the white sides to the black sides (clockwise in the diagram viewed from above). However, the vanes are observed to rotate in the opposite direction. It can be shown by a simple estimate that the radiation pressure is extremely small here.

The observed behavior is due to a thermal effect associated with the gas in the bulb, and can be overcome only with a very high vacuum and a pivot with a very small amount of friction.

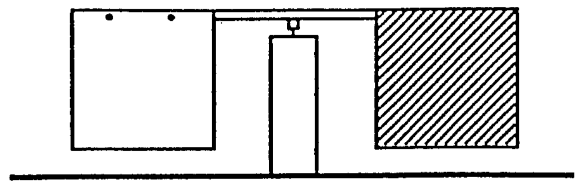
It is a common misconception that the behavior can be explained as follows. The volume inside the glass bulb is not a high vacuum, and may not be evacuated at all. When the radiometer is exposed to light, the black side becomes hotter, which increases the temperature and thus pressure of the gas on this side. This increased pressure causes the rotation from the black sides to white sides. In addition, close inspection reveals that the vanes typically have a slight twist about a horizontal axis such that the upward convection near the black sides may help drive the assembly. This argument is incorrect because any pressure imbalance between the gas near the black and white sides quickly equilibrates at the speed of sound. Although the average speed of the gas particles is greater near the black side because the temperature is greater, the density is less such that the pressures are the same.

An understanding of the behavior of Crooke's radiometer occupied some of the most capable scientists of the late 1800s, including Maxwell, Stokes, and Reynolds. The following physical explanation was given by Einstein in 1924. Consider gas within roughly a mean-free-path length from the edge of a vane (refer to the diagram below). A gas particle moving toward the black side has an average speed dictated by the greater temperature on that side. The particle is impeded not only by similar particles from the black side but also by particles from the white side. Because the latter have less average speed, the particle is impeded *less* than it would have been if it were nearer the center of the black side (i.e., several mean-free-path lengths from the edge). The particle thus strikes near the edge of the black side with greater momentum. Similarly, a particle on the white side strikes near the edge of that side with less momentum.



Explanation for the behavior of the radiometer. A gas particle striking near the edge of the black side is impeded less because the particles from the white side are at less temperature. Similarly, a particle striking the white side is impeded more.

SO.11b Acoustic radiometer



Acoustic radiometer, showing the metal side of the left pane and the foam rubber side of the right pane. When placed in a reverberating acoustic enclosure (not shown) ensonified with noise, the panes rotate clockwise viewed from above.

When any type of wave is incident on a surface, there can be a nonzero time-averaged pressure, or *radiation pressure*. For an absorbing surface, this pressure is less than that for a reflective surface. Hence, when two such surfaces are joined back-to-back and placed in an isotropic and homogeneous wave field, there is a net force in the direction of the reflective to the absorptive side. If this device is mounted such that it is allowed to spin as a result of the net force, the system can be considered a *radiometer* because it detects the presence of the wave field. However, the well known electromagnetic radiometer spins in the *opposite* direction (counterclockwise viewed from above) when exposed to light of sufficient intensity (refer to SO.11a).

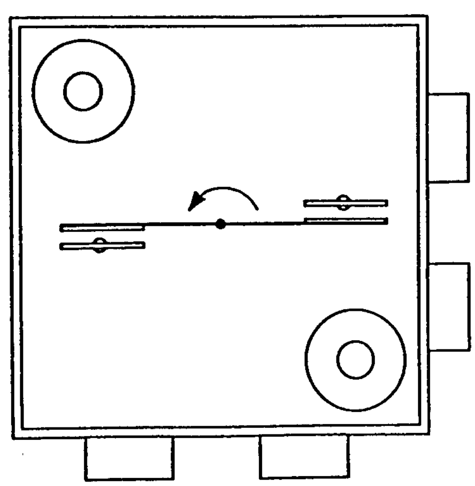
The acoustic radiometer shown above can be used to demonstrate radiation pressure.[1] Each pane has an aluminum side and a foam rubber side. The panes are attached to a rod that is supported by a pivot that has with very little friction. When placed in an enclosure and exposed to isotropic and homogeneous acoustic noise of sufficient intensity, the apparatus spins in the direction predicted by radiation pressure (clockwise viewed from above).

Preliminary quantitative results show that the actual torque is at least 50% greater than that predicted by theory. Research is currently being conducted to confirm and understand this.

[1]Timothy G. Simmons, Bruce Denardo, Andrés Larraza, and Robert Keolian, "Acoustic Radiometer Demonstration," *Proc. of the 16th International Congress on Acoustics and 135th Meeting of Acoustical Society of America*, 20-26 June 1998, Vol. I, edited by Patricia Kuhl and Lawrence Crum, pp. 129-130.

SO.11c Acoustic Casimir effect

Top view of vertical metal plates, where one of each pair is fixed and the other is attached to a rotatable arm. The system is enclosed in a clear acrylic box with compression drivers attached to the bottom. The plates are attracted when high-intensity acoustic noise is present.



The acoustic enclosure in previous demonstrations is utilized here, where two compression drivers are employed to generate high-intensity noise. Metal plates are vertically attached to the ends of an arm that is pivoted at its center (refer to the diagram above). Two other plates are fixed such that they are parallel to and near the movable plates. When the drivers produce high-frequency (e.g., 7.5 kHz to 15 kHz) noise of sufficient intensity, the arm rotates counterclockwise in the above diagram; that is, the plates are attracted.[1,2]

This effect is due to an imbalance in radiation pressure. The noise is approximately uniform and isotropic outside the plates, and is composed of nearly a continuum of modes. Between the plates, however, the boundary conditions due to the plates reduce the number of modes and thus the noise intensity. In fact, if the distance between the plates is less than the smallest half-wavelength of the noise (at the upper limit of the band) the noise intensity between the plates is zero. The reduction in intensity between the plates causes an imbalance in radiation pressure, so the plates are attracted.

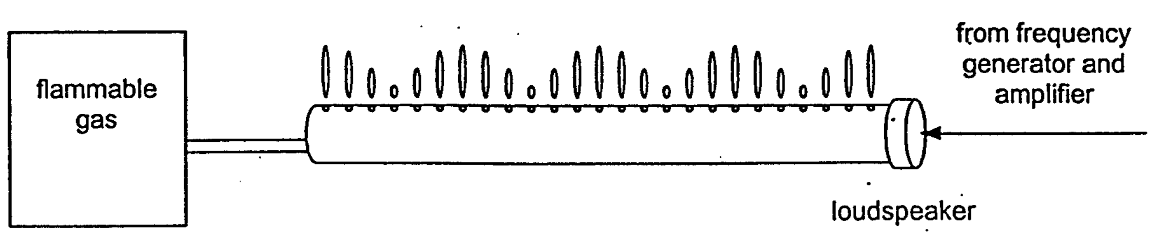
An analogous imbalance in radiation pressure is responsible for the celebrated *Casimir effect*, in which two parallel conducting plates in a vacuum are slightly attracted due to the existence of the zero-point quantum electrodynamic field (virtual photons). The boundary conditions reduce the intensity of the field between the plates, which results in attraction.

It is remarkable that, under certain conditions in the acoustic case, the force can be *repulsive* instead of attractive. This effect is not due to resonance amplification of the modes between the plates, which is negligible here because the plates are open on the sides. The repulsion is a result of a nonzero lower limit of the band of noise.[1,2] This effect does not occur in the Casimir case for parallel plates because the lower frequency limit of the zero-point field is zero.

[1]Andrés Larraza, "A demonstration apparatus for an acoustic analog to the Casimir effect," *American Journal of Physics*, vol. 67, pp. 1028-1030 (1999).

[2]Andrés Larraza and Bruce Denardo, "An acoustic Casimir effect," *Physics Letters A*, vol. 248, pp. 151-155 (1998). Andrés Larraza, Christopher D. Holmes, Robert T. Susbilla, and Bruce Denardo, "The force between two parallel rigid plates due to the radiation pressure of broadband noise: An acoustic Casimir effect," *Journal of the Acoustical Society of America*, vol. 103, pp. 2267-2272 (1998).

SO.11d Flame tube

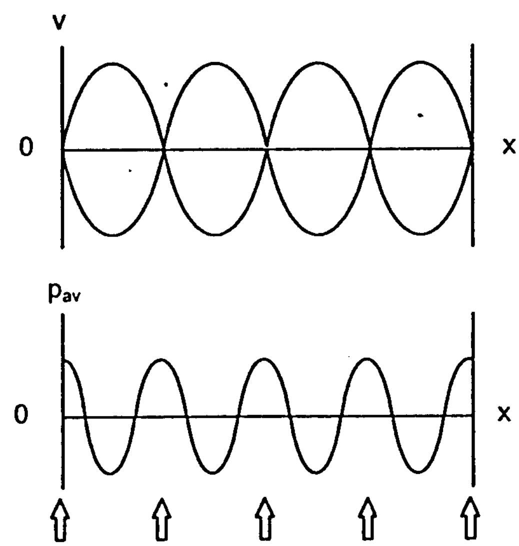


In the flame tube, the velocity distribution of a standing wave is shown by the variation in the heights of flames from small holes in the tube.

In this very dramatic demonstration, propane or methane flows into a closed-end metal tube that has small holes drilled along its length. A loudspeaker at one end of the tube is driven with a frequency generator and amplifier. At resonance frequencies, the heights of the flames vary along the length and are maximal at velocity antinodes of the standing wave. Music (a superposition of frequencies) can then played through the loudspeaker, creating a dramatic effect. The finale of Stravinsky's *Firebird* is both effective and appropriate for this purpose.

The effect may be due to radiation pressure, which is the time-averaged pressure. The radiation pressure of a wave equals the average potential energy density minus the average kinetic energy density.[1] Note that this is in accord with Bernoulli's law for incompressible flow, which yields a reduction in pressure at points of increased velocity. The radiation pressure of a standing wave in a closed tube varies as in the diagram below. The greater pressure at the pressure antinodes (velocity nodes) could be responsible for the flames being higher at those points.

Peak velocity and radiation pressure corresponding to the standing wave in the above diagram. The maxima (arrows) of the radiation pressure distribution could be responsible for the higher flames at those points.



A possible problem with this explanation is that it may not account for the surprising fact that some flame tubes have maximum flame heights at the *opposite* locations (pressure nodes or velocity antinodes).[2]

[1]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon, New York, 1987), pp. 257-258. For a closed tube, the radiation pressure can include a spatial constant in addition to the energy density terms, but this does not appear to have any effect here.

[2]Harry F. Meiners, ed., *Physics Demonstration Experiments*, Vol. I (Ronald Press, New York, 1970), pp. 495-496.

NOISE AND SENSORS

Thomas G. Gabrielson
Pennsylvania State University

INTRODUCTION

DR. GABRIELSON: The last few times I have been to PASS I have spoken about the physics of sensors in relationship to self-noise and in relationship to signal response. This year, I'm not going to speak about signal response. Instead, I'm going to expand the material on sensor noise because it is truly fascinating in its own right.

[Transparency I-1]

A lot of people believe that any time I talk, the talk itself is noise, but I had an illuminating conversation with Dr. Hargrove at the beginning of PASS this year. He pointed out that there is a critical difference between noise and lack of information. For example, my talks are lack of information, not noise. Noise contains a great deal of information and I hope that by the end of the lecture, I will have convinced you of that.

First, where did my interest in sensor noise originate? A number of years ago, we started examining some very high-performance sensors of various sorts; sensors intended for detection of very small signals of different kinds.

For example, inertial navigation systems have accelerometers and angular-rate sensors (gyroscopes). For long-term inertial navigation you have to detect very small signals and sensor noise becomes very important as a limit to the ultimate performance.

At about the same time, I got involved in microfabrication (MEMS) technologies, and there noise is also critically important. Normally, noise becomes more significant when a sensor or device is made smaller. So I started studying noise and it has really been fascinating. It has been a good education for me; hopefully, you will get something out of it, too. If not, you can always write nasty things on the evaluation.

[Transparency I-2]

Where are we going to go and what are we going to do? I am going to break this talk into three parts. First, I am going to talk about mechanical thermal noise; that is, the noise associated with molecular motion, the normal molecular motion that exists at the temperature at which we operate our system.

I am not going to talk about very low-temperature systems or superconducting sensors. I am not going to talk about sensors that are so small that they support only a very small number of thermally induced modes. These are interesting cases but generally less important from a practical standpoint and they would make the talk a little bit too long. Even sensors that operate at cryogenic temperatures generally show fully developed thermal fluctuations so these exclusions will not limit the applicability much.

Second, I will discuss shot noise. This is noise associated with movement of discrete independent particles of some kind. They could be photons, they could be electrons, or they could be particles of sand dropping on a surface. This is a good place to introduce some other concepts as well.

Finally, in the third hour, I will pick some topics that are favorites of mine and talk them. This part of the lecture will be a little bit more disconnected but you will see the roots of the earlier discussions in these topics.

I. THERMAL NOISE

[Transparencies I-3 and I-4]

If I consider a very simple system – like this demo, a mass on a spring – and I displace it a bit then let it go, it vibrates. If the system is attached to motionless support (and not my hand!), the oscillation will decay to any arbitrarily small amplitude if you wait long enough, right?

Let's consider a system that has a resonance frequency of 100 Hz, a Q of 10, and a mass of a milligram. It is not outrageous to think of a sensor built with these parameters. If you displace the mass by a millimeter, how long does it take for the amplitude to decay to 10^{-8} mm? Actually, it does not take very long. You can calculate the exponential decay rates in the simple manner of Q-decay of oscillations. Is 10^{-8} mm a reasonable amplitude? If this system were an accelerometer, that amplitude would correspond to an applied acceleration of just under a micro-g, and there are certainly plenty of times we want to be able to sense a micro-g at 100 Hz.

How long does it take to decay to 10^{-8} mm? *At room temperature it takes forever*; it never reaches that level because of thermal fluctuations. Molecules in this system are fluctuating and the system is coupled to all kinds of other fluctuating things. That is the subject of the first hour: simple thermal fluctuations.

[Transparency I-5]

Most of the time when we are taught simple harmonic oscillators, we are shown a mass on a spring, and we study that for a while, and then throw in a damping element and study that for a while. You consider an initial displacement. Or, the mass might have been given an initial velocity instead of an initial displacement. It decays to some arbitrarily small level according to this homogeneous differential equation on the left. You all have seen this, hopefully: mass times acceleration plus resistance times velocity plus the spring constant times the displacement with the sum set equal to zero.

But this is really not correct. This is bad physics. What really happens is that the mass motion decays into a background of noise. Many times in experimental work you are so plagued with things like 60-Hz interference (or 50-Hz interference, depending on which side of the ocean you work) that you forget about the fundamental noise processes associated with these sorts of motions. There is a fundamental fluctuating force that is directly related to the introduction of damping in this system.

On the right is the proper equation of motion for a damped system and above it is the proper diagram for that system. If you put in a damper, that fluctuating force generator must go in, also. It's bad physics to leave out the force fluctuations.

[Transparency I-6]

A number of classic papers have been written on this subject. One classic theorem is the fluctuation-dissipation theorem. This theorem can be expressed in terms of an equation, but, for our discussion, the equation is not too important. The idea behind the theorem is more important so I am just going to talk you through the theorem.

If we look at a sensor like a condenser microphone – a structure that has a fixed back plate and a movable diaphragm – we can identify various damping mechanisms. The volume between the diaphragm and the back plate is filled with gas and there is a viscous damping associated with squeezing this layer of gas in and out as the diaphragm moves toward and away from the back plate. Also, the diaphragm and its edge support have mechanical loss. And, the diaphragm radiates sound, which also causes energy to leave the system.

Normally, radiation is not called damping; however, any path by which the energy can leave the system is, from the system's point of view, a loss of energy. From the perspective of the diaphragm, radiation is a loss mechanism.

Any time there is a path from the system out to the environment along which energy is lost, then that same path also permits energy exchange from the environment back into the system. Because the path is bi-directional, the fluctuations in the environment that result from normal molecular motion act back on the system and keep the system from decaying down to any arbitrarily small level.

[Transparency I-7]

Frequently you see this sort of phenomenon explained in terms of the principle of equipartition. Equipartition describes the total thermal energy in the system. Any system may have many different modes of vibration. It may have a mode characterized by twisting back and forth – a torsional mode. It may have a rocking mode. It may have a breathing mode or asymmetric elastic modes of vibration.

For any one of these modes, the total energy from thermal agitation is ½ times Boltzmann's constant, k_B, times the absolute temperature of the system. For a simple mass on a spring, there is kinetic energy associated with the velocity of the mass. There will be a thermal energy of ½ k_BT associated with that mass motion. There will also be potential energy associated with the compression or extension of the spring. A thermal energy of ½ k_BT will be associated with the elastic energy storage in the spring. The total energy for the simple vibration is then k_BT.

What about the value of Boltzmann's constant? It is a pretty small number, so you think, well, what is the big deal: k_B times T at room temperature is about 4 times 10^{-21} joules. It sounds so small how could there possibly be any influence on a system? But intuition is misleading here. In sensors, we often can measure very small quantities and that seemingly small thermal energy can overwhelm a sensitive measurement.

If you consider a molecule in a liquid, each independent direction of translation has a thermal energy of ½ k_BT. This is a statistical average. All of the molecules do not have same energy. The x component has an average energy of ½ k_BT, the y component has an average energy of ½ k_BT, and the z component has an average energy of ½ k_BT as well.

But this is not only true for individual molecules. It is also true for macroscopic objects like a ball bearing in a liquid. As for the molecule, for each independent direction of translation, the thermal component of the kinetic energy, ½ mv^2, is equal to ½ k_BT. What is the difference? The mass is much larger for the ball bearing than for the molecule, so the velocity is considerably smaller. Even so, the effects can still be strong enough to limit our ability to measure the ball bearing's position (for example).

Other sorts of systems have a corresponding store of thermal energy. In a capacitor, you would write the stored energy as one-half times the capacitance times the voltage-squared. The equilibrium energy of fluctuation that results from thermal agitation is $\frac{1}{2} CV^2$ equal to $\frac{1}{2} k_BT$. In any case where we can write the energy as some scalar times some dynamic quantity squared, there will be a component of thermal energy in equilibrium with a value of $\frac{1}{2} k_BT$

[Transparency I-8]

I would like to make a point about molecular agitation with an example – a protozoan [*Loxodes striatus*]. This is an interesting little creature. It is 200 microns long. It has, up at the top in this picture, a very small chamber and I've also shown an expanded view of that area. Biologists believe, in fact, that this animal exists only as a male, there are no females of this variety, because the head area is so small. (*laughter*)

The chamber that I've highlighted is about 7 microns in diameter. In that chamber is a little nodule – a mineral accretion on the end of a hair-like structure. If the animal is oriented as shown in the photo, the hair bends away from the "head"; if the animal turns over, the hair bends toward the head. This bending creates a chemical interaction at the root of the hair structure and this chemical signal adjusts the way the flagella around the organism move.

If the organism is pointed in one direction with respect to the gravity vector, then the flagella act in coordination and the organism travels in more or less of a straight line. If it is pointed in the other direction, the organism swims randomly.

Why would it do this? That is a fascinating story and I can only give you the highlights. It likes to find an optimum dissolved oxygen concentration in the fluid in which it lives. It uses a chemical sensor to figure out what the concentration is and it uses the gravity sensor to migrate toward it or away from it, depending on what the concentration changes are. Stratification occurs in lakes and ponds, so we know the oxygen concentration varies most strongly in the vertical direction; consequently, a gravity sensor is useful.

Let's consider the sensor's potential energy. I will just estimate it as the mass (about 45 picograms) times the gravitational acceleration (9.8 m/s^2) times the range of the mass motion (about 3 micrometers). Now the ratio of that potential energy to the total thermal energy, kT, is about 350, a reasonably large number, so there is enough change between the up and down orientations that this animal can sense gravity unambiguously.

The animal is not fixed in any particular orientation; it is not sitting motionless for minutes at a time, so it cannot do much integration. It is moving around all the time, so it does not have much integration time to improve the estimate. If you reduced all of the dimensions linearly by a factor of four, then the ratio of available potential energy change to thermal energy would be about one and the organism would not know up from down.

This is a mature adult. A juvenile might be about half that size or so. There is probably some excess sensitivity in the adult so that the juvenile with its smaller sensor can still survive. But, all things considered, nature has contrived to push this sensor pretty close to the edge of the thermal limit.

When I talk about high-performance sensors, I don't necessarily mean detection of extremely small signals. This protozoan has a very high-performance sensor, because it's performance is excellent *for its size* – this sensor is extremely small. Not only that, it reproduces, which is unusual for a sensor. (*laughter*)

So high performance might mean extremely small size. Or, it might mean a device that is extremely cheap – our sponsors like that sort of argument as well. All of these factors enter into the design of sensors. But, I cannot take credit for the design of the protozoan.

DR. MAYNARD: How close is the human ear to the kT limit?

DR. GABRIELSON: I have read different accounts, but the majority claim that a young person's ear – a pre-Sony-Walkman ear – is very close to the thermal limit. But, I've never tried to investigate the threshold myself so I honestly don't know. My own ears have degraded substantially so I won't be the test subject.

DR. MIGLIORI: Is the cavity in the protozoan an open space or is it fluid filled?

DR. GABRIELSON: It is fluid filled.

DR. MIGLIORI: So it is damped.

DR. GABRIELSON: Yes, strongly.

DR. MIGLIORI: So that kills your integration time. That is why you say there is no integration?

DR. GABRIELSON: Yes, that is a good point. The damping and the continuous motion of the organism both conspire to prevent effective signal integration.

[Transparency I-9]

A free body in a bath of thermally agitated molecules is affected by random collisions. The smaller the body, the larger the effect. For example, the influence of thermal agitation on the small nodule in the gravity sensor of the protozoan is much greater than the influence of the molecular collisions on the protozoan itself.

Einstein studied the process of molecular agitation of free particles. As he showed, the root-mean-square displacement is proportional to the square root of two times the diffusion coefficient times the time. As time increases, the particle "wanders" further from its initial point as the square root of time. The diffusion coefficient is kT divided by the equivalent mechanical resistance.

If we consider the RMS drift in one second for various size bodies, we can see the strong dependence on size. In the table shown, the bodies are considered immersed in water (and considered to have the same density as water). A body with a radius of one meter is virtually unaffected and a body with a radius of one millimeter shows very little influence. When the radius shrinks to one micrometer, the drift is about equal to the size of the body; smaller particles are mercilessly beaten about by molecular collisions. This has obvious relevance in the design of microscale "robots" for fluid environments. In paper designs, the problems of stability in the environment of molecular agitation are often ignored.

[Transparency I-10]

Consideration of the total thermal energy is educational and we can draw some interesting conclusions from it, but we will get a lot more mileage by establishing the frequency distribution of the noise. Fortunately, Nyquist solved that problem. He gave us a very simple procedure for finding the frequency distribution.

Nyquist's expression is remarkably simple. If we take the mechanical resistance (that is, the ratio of force to in-phase velocity) then multiply by four times Boltzmann's constant times the absolute temperature times the relevant interval of frequency, you get the mean-square fluctuation force associated with that interval of frequency. It is very simple.

Johnson noise in electrical resistors is written the same way. The mean-square fluctuation in voltage is equal to the electrical resistance times $4k_BT$ times the interval of frequency. (Quite often we divide both sides of this equation by the frequency interval to produce the spectral density of the fluctuations on the left-hand side.)

Now, this is not quite right. If you integrated this expression over all frequency, you would get infinite power. The form of Nyquist's relation that is generally valid at all frequencies is also included on the slide. I wanted to show you the exact form of equation so you would see that it really does behave properly. This form is a quantum-mechanical necessity. However, the complicated factor is unity for every problem that we will do.

If the frequency is high enough or the temperature is low enough, then the correction does become important and you need to know that it is there. What the extra factor does is to roll off the noise contribution rapidly after reaching the point where Plank's constant times the frequency equals Boltzmann's constant times the temperature (that is, where the thermal energy equals a quantum of energy).

At 1 kelvin you would need to go up to about 2 GHz to reach the break point. At one millikelvin you would still have to go up to 2 MHz to reach the break point. It is not out of the question to have to cope with this roll off but in most practical sensors, the roll off is not an issue. Please remember that the roll off exists; you never know where your research will take you.

At more common temperatures and frequencies, the simpler expression applies. We can use this relationship to estimate Boltzmann's constant if we believe the voltage measurements that we make and we have confidence in our knowledge of resistance and temperature. All it takes is a good, low-noise amplifier and some care in the measurement. In principle, we could also estimate Plank's constant by carrying out the measurement at sufficiently low temperature and sufficiently high frequency but this would be much harder. A bit later in the lecture, we will see how to estimate the elemental charge on an electron from a noise measurement. "We don't need no stinkin' signals" to do physics. (*laughter*)

[Transparency I-11]

Having said that, I will immediately qualify my position. We should not get too narrowly focused on noise. What is really important in most systems, is the signal-to-noise ratio. We would like to get as high a signal-to-noise ratio as possible for a given cost or a given size or for some other constraints.

There are many ways of tracking the signal-to-noise ratio in an analysis. Instead of talking about a fundamental quantity like the noise voltage, we can translate that quantity into the effective signal level that would produce that noise voltage. This is the signal-equivalent noise. By doing this, you can avoid being misled about the achievable signal-to-noise.

Let's consider a simple accelerometer, a mass-spring system inside a case. If the case moves up and down, a relative displacement is produced between the mass and the case. If we set the noise generator to zero, we can calculate the signal response. Then we set the signal (the applied acceleration) to zero, which we could do by clamping the case to an immovable object, and solve for the response to the noise generator. From those two values – the zero-noise signal response and the zero-signal noise response – you can calculate the noise-equivalent signal. For example, if the voltage-noise level of an accelerometer is 1 microvolt RMS (root-mean-square) and the signal response is 0.1 volt per g, you would say that there was a 10 micro-g RMS noise-equivalent signal. The noise-equivalent-signal specification is far more useful because it connects the limiting performance of the accelerometer with the physical quantity that you are interested in measuring. Just having a very small noise is not in itself valuable; however, if that corresponds to a very small signal, then you have something.

[Transparency I-12]

Let's consider the accelerometer. If calculate the noise of the system and the response of the system, we discover that the response of the mechanical system is not flat with frequency – it has a resonance. When we calculate the noise-equivalent acceleration, though, we find that it is independent of frequency because both the signal and the noise response have the same spectral shape.

The result is rather simple. The square of the mass times the noise-equivalent acceleration is equal to the Nyquist value for the square of the fluctuating force. Solving for the spectral density of the noise-equivalent acceleration (in meters per second-squared per root hertz), we obtain the expression, $4kT$ times the mechanical resistance divided by the mass squared. Alternatively, we can express the result in terms of the system Q. In that form, the noise-equivalent acceleration is equal to $4kT$ times the resonance frequency, ω_0, divided by the product of the mass and the Q.

How can we lower the noise-equivalent acceleration for this simple system? We can increase the mass. If the damping, R, is fixed, then the noise-equivalent acceleration is inversely proportional to the mass squared. In changing the system, R may also change, so you must be careful to consider all of the implications of a structural modification. However, in any case, the mass is in the denominator so, if the noise from thermal agitation is the limiting factor, increasing the mass lowers the noise-equivalent acceleration.

One of the popular new technologies is microfabrication or MEMS (Micro-Electro-Mechanical Systems). By microfabrication, we can make very small structures that can perform the same function as larger, conventionally machined devices. But, if we try to make a very small accelerometer, we must contend with the fundamental problem associated with a very small moving mass – the noise-equivalent acceleration that results from thermal agitation can be very high. Sometimes it is implied that the only reason we have not made extremely small, highly sensitive sensors in the past is that we do not have suitable machining processes. It's not simply a matter of miniaturization though. By making the sensor physically smaller, we are increasingly subject to molecular agitation and if we are unaware of such processes, we can make unrealistic, unsupportable claims.

By intelligent design, however, we can push the performance as far as nature permits. For example, if we decrease the resonance frequency (and can accept the lower bandwidth), the noise-equivalent acceleration can be reduced. If we increase the Q, we can also reduce the noise-equivalent acceleration. Both of these changes have implications in performance but, with some thought, we may be able to make the consequences acceptable.

Bruel and Kjaer make a low-noise microphone that embodies this design philosophy. In this microphone, they make the mechanical Q considerably higher than in a conventional condenser microphone. This adds a significant peak in the spectral response at the high-frequency end. Since this departure from flat response is not desirable, they electronically compensate the response to flatten it. The result is a microphone with lower self-noise than conventional designs of the same dimensions.

Once you understand the basics, a bit of unconventional thinking can produce sensors with higher performance than obtainable with standard approaches. This is the sort of philosophy that I hope you will make your own this week: try to reach a deep understanding of the fundamentals, then apply that understanding to your research instead of working blindly through the list of "tried-and-true" methods.

These concepts of thermally induced fluctuations are intimately tied to the thermodynamics of systems in equilibrium. One of the nice things about equilibrium thermodynamics is that we can solve a problem in different ways. Even if we use very different techniques, we wind up with the same answer. Now, I would like to show you a nice example of that.

[Transparency I-13]

Let's consider the noise associated acoustic radiation. Remember that, when I was introducing the fluctuation-dissipation theorem, I told you that radiation from a system counts as a loss mechanism. Radiation provides a path by which energy can leave the system, so we ought to be able to calculate what the corresponding noise is.

For example, consider a small source – a monopole – that generates spherical waves. The source could either be very small and of arbitrary shape or it could be a spherical source operating in the spherically symmetric "breathing" mode. The spherical wave solution has the form of a radially expanding traveling wave with an amplitude inversely proportional to the radial distance, r.

The volumetric-density form of Newton's law relates the gradient of pressure (force per unit volume) to the density and the time derivative of velocity. If we rewrite that equation using complex exponential forms for time dependence and evaluate the gradient in the radial direction, we can solve for the radial particle velocity, u_r.

Once we have expressions for both pressure and radial velocity, we can write the impedance. The mechanical impedance is the pressure times the area divided by the velocity.

For simplicity, we can let the product of wave number, k, and radius be very small (i.e., the source is much smaller than a wavelength). This is not necessary but it makes the algebra easier. Having done that, the real part of the impedance approaches the product of density times sound speed times area times the quantity $(kr)^2$. This we can rewrite as pi times density times frequency squared times area squared divided by sound speed. Now that we have the real part of the radiation impedance, we can immediately write the spectral density of the noise-equivalent pressure. The noise-equivalent pressure squared is the noise-equivalent force squared divided by the area squared. The result is that the noise-equivalent pressure squared is $4kT$ times π times ρ times f^2 divided by c times whatever bandwidth we are considering. This is a nice, relatively simple expression.

[Transparency I-14]

At this point, we've solved the problem, but let's try to solve it in an entirely different way. Let's start out by putting a very small pressure sensor in a box – a cube with perfectly rigid walls.

The sensor is located at the center point of the box. The box dimensions are $2L$ by $2L$ by $2L$; the shortest distance from the source to a wall is L.

What are the acoustic modes if this is a rigid-walled box? The modes are standing waves having pressure maxima at the walls. From symmetry, these modes either have pressure maxima or pressure zeros at the center. Any mode having a pressure zero at the center is not sensed by our detector so we will not include those modes in our solution.

For maxima at the center (0, 0, 0 in our coordinate system), the relevant modes will be cosines. The arguments for those cosine functions that produce maxima at the walls are given by any integer times π times x divided by L. The component of the wave number in the x direction is just the factor multiplying x, so the wave number component is always an integer times π divided by L for any of the three coordinate directions.

What is the spacing between these allowed wave number components? It is just π/L. The total wave number squared, k^2, is the sum of the squares of the wave number components. If we plot the locations of all of the allowed wave number components in wave-number space, the result is a very simple cubic lattice with evenly spaced points.

The dimension of a unit-cell in k-space is π/L. Consequently, the unit-cell volume is $(\pi/L)^3$. Also, there is one k-point per cell. (If you locate the eight k-points that bound a unit cell, an eighth of each is inside the cell, so you really have one point per unit-cell.)

By equipartition, each mode has $\frac{1}{2}kT$ associated with the kinetic energy and $\frac{1}{2}kT$ associated with the potential energy, so the total equipartition thermal energy per mode is kT. Furthermore, the density of thermal energy in k-space is the energy per wave number (i.e., per mode) divided by the volume (in k-space) per wave number.

[Transparency I-15]

In principle, the problem is solved; however, we are in k-space and we really want to be in frequency space. So we need to figure out the relationship between dk and df. What is dk? It is the region in k-space between k and $k+dk$, which is a spherical shell. The volume of that spherical shell is $4\pi k^2\, dk$. The energy, dE, in the shell is the energy density times the volume of the shell. At this point, we let the box dimension, L, go to infinity so that there is a continuum of modes and we can consider the differentials as differentials of a continuous function rather than continuing to count individual modes.

The relationship between wave number and frequency gives the corresponding relationship between *dk* and *df* so the energy increment, *dE*, can be written in terms of *df*. If we divide by the spatial volume of the cube ($2L^3$), we obtain the energy density for the increment of frequency, *df*. But we also know that the energy density can be written as pressure squared divided by the quantity ρc^2. Therefore, we set these quantities equal and solve for the pressure squared. The result is the same as before even though we used a radically different approach. The noise-equivalent pressure is the same either way and that is very satisfying. Of course it could mean that I have made exactly the same mistakes in both problems. (*laughter*)

DR. MIGLIORI: Do you want to comment about the fact that your system has an infinite number of modes, whereas most systems do not?

DR. GABRIELSON: Very good point. Thank you for bringing that up. My ideal box is infinitely large so the number of modes is infinite and they are arbitrarily close together in wave-number space. This is equivalent to the radiation-impedance solution since those results were derived under the assumption of propagation out into infinite space. So the two idealized situations that I considered are compatible.

In a real system, the boundaries are not at infinity and there are a finite number of modes. If they are spaced so closely and have sufficient damping that their response functions overlap to form a virtual continuum, we don't need to make any adjustments to what we've done above. Often, however, there are distinct modes with sufficient separation that the continuum approximation is not reasonable. In this case, equipartition still applies but the thermal energy is "concentrated" in the individual modes and the spectral distribution depends on the modal damping (or Q). Experiment and theory still agree but we have to be more careful with the calculations and drop the continuum assumption.

[Transparency I-16]

One simple adjustment that we can make is to relax the assumption that the sensor is small with respect to a wavelength. In the case of a spherically symmetric body or a circular piston in an infinite rigid baffle, the expressions for the real part of the radiation impedance are not particularly complicated regardless of the frequency. For our "small" sensor approximation, the noise-equivalent pressure was directly proportional to frequency. For both the sphere and the piston, the noise-equivalent pressure becomes independent of frequency when size of the sensor exceeds the acoustic wavelength. The modal structure in the infinite-box model has not changed; however, when the sensor's active surface becomes large enough, the mode wave function is averaged over the active surface and the contribution of the higher-frequency modes is reduced by this area averaging. It is unnecessary to treat the details of the individual mode shapes, though. We have an expression for the real part of the radiation impedance and Nyquist's theorem gives us the correct answer without working nearly as hard as we would if we continued the modal approach.

[Transparency I-17]

The next topic I'd like to address is the connection between noise and causality. On the surface, this seems like a strange association but there is an incredible richness in the study of fluctuations and a number of fascinating relationships between "random" aspects and "deterministic" aspects of real systems. Nyquist gave us the first such example in the intimate connection between the real part of an impedance and thermal noise. Kramers and Kronig give us the next.

If we examine a linear system by applying some forcing function, we observe some response. We can characterize the system as some function $h(t)$, given an input f_{in}, and a corresponding output f_{out}. One way of writing a general relationship between the input and the output is to write two terms: the first is merely a constant times the input (an instantaneous response); the second is a convolution integral (a response that may extend beyond the input in time).

Jackson uses this model for the general, linear system response. It's a clever formulation but it did not strike me how clever it was until I started doing some calculations. The separate term for the instantaneous response is unnecessary. We could, instead, drop the first term and include a delta function in the definition of $h(t)$. However, the instantaneous response occurs frequently enough that it's useful to consider it separately and dispense with the delta functions. For example, if the system is a resistive voltage divider, the transfer function is just a real constant. Simple delayed responses can also be treated as separate terms. As long as there is a part of the system that acts so as to reproduce the input waveform with only an amplitude scaling or a shift in time, treating that part as a separate term simplifies the analysis.

In a physical system, we do not want the output to anticipate the input. If there's an output before the input starts, then something's wrong. Consequently, we have a condition on the function, *h(t)*. The function, *h*, must be zero if its argument is less than zero. This simple condition leads to some interesting consequences.

[Transparency I-18]

Consider the Fourier transform pair of *h(t)* and *H(ω)*. In the transform from the time domain to the frequency domain, we do not have to integrate from negative infinity to infinity because the integrand is zero for negative time.

Now, the Fourier transform implies that we are examining the function along the real-frequency axis, but we can continue the solution into the complex *ω*-plane. In the transform from the frequency domain into the time domain, we can allow *ω* to be complex. If we consider complex frequency, the exponential factor becomes the combination of the usual oscillatory factor and an exponential-decay (or growth) factor.

If the imaginary part of *ω* is negative, then the complex exponential factor grows with *ω*. In order for the result to be bounded, *H(ω)* must be well behaved in this region. In contrast, if the imaginary part of *ω* is positive, then the exponential decay of the complex exponential factor in the integrand can work against singularities in *H* so that the integral is still well behaved. So for all the physically realizable *h(t)*'s we encounter, *H* must be analytic if the real part of frequency is less than zero. To say this another way: if a system is causal (that is, its output does not anticipate its input), then *H(ω)*, the Fourier transform of the system response function, must be analytic in the region where the imaginary part of *ω* is negative.

Furthermore, we need to examine the behavior of the integrand when *ω* is real since that's the path of integration for the inverse transform integral. We certainly want the integrand to go to zero as the frequency goes to either positive or negative infinity. We will insist on that.

We will also require the integrand to be finite along the real axis. We will not allow poles for real frequency. In other words, we are not analyzing ideal, dissipation-free systems. In all of our real systems, we have dissipation even if it's small so the poles are never right on the real axis. Neither of these two additional conditions is an obstacle to understanding physical systems. There will always be idealized, unrealizable systems that can cause problems but we are not interested in abstractions.

[Transparency I-19]

Now, for integration in the complex plane, we look for poles, among other things. If the contour of integration surrounds a pole, the contribution to the integral associated with that pole is the residue of the integral at that point. The residue is related to the integrand by an imaginary factor. What happens when you multiply a complex number by the simplest imaginary number, j? The real parts become imaginary parts and the imaginary parts become real parts. So, if we introduce a pole by dividing the function, $H(\omega)$, by $\omega - \omega_0$, then integrate along the right contour, we can swap real and imaginary parts. Such an integral is shown here. This is, in fact, a Hilbert transform, which Dr. Waxler pointed out to me last week.

If we introduce the simple pole right on the real axis, and perform the integration along the real-ω axis, then we can construct a closed-loop path as shown. The arc in the lower half-plane closes the path. Because of our previous constraints on $H(\omega)$, the integral over this arc path is zero. Therefore, the only contribution to the closed-path integration is from the residue on the real axis. (And because of the path deformation, we only use half the residue.)

There is another clever trick that can make evaluation of the integrals easier. The integral from minus infinity to infinity of any constant over a simple pole is zero. That's because it's an odd function integrated over limits symmetric about zero. This maneuver is unnecessary from the point of view of the theory but it can make numerical evaluation much easier since the integrand is much better behaved near the pole. Most numerical integration codes will not sample exactly at the pole so the apparent problem of zero over zero does not come up. Even if the code does sample right at that point, a little manipulation of the limits is usually sufficient to shift the sampling away from the pole.

[Transparency I-20]

This leads us to one of the classic forms of the Kramers-Kronig relationships. In science, we never call something by a descriptive name; we always use some person's name and we do that, because if we called everything by a descriptive name, you would not need to go to graduate school to understand things: everything would be obvious. (*laughter*)

To write the Kramers-Kronig relationships we take the result from the last line on the last slide and separate the real and imaginary parts. The imaginary part is equal to an integration involving only the real part and the real part is equal to an integration involving only the imaginary part. The consequence is that the real and imaginary parts of a causal function are related!

Impedance is one of the functions to which we can apply the constraint of causality. The impedance relates the current produced to the voltage applied and, in a physical system, we can't have a current produced before the voltage is applied.

Now this is not only intellectually interesting, this has implications for the noise of a system. The thermal-agitation noise is solely a function of the real part of the impedance but the Kramers-Kronig relationships say that we could determine the noise even from the imaginary part. (Or, by measuring the noise, we could calculate *both* the real and imaginary parts of the impedance.) In order to complete the integrals, we need to know the behavior over all frequencies but, many times, the integrand goes to zero rapidly enough that we can evaluate the integrals with only a limited knowledge of the integrand. So, in principle, we can compute the total impedance from a noise measurement. I made this slide before I tried it, so I wrote "usually easier said than done." That is always a safe statement to make when you're talking about a measurement.

[Transparency I-21]

In spite of the risk, I decided to try the measurement and the analysis anyway. I examined a very simple system – the parallel combination of a resistor and a capacitor. I used values that I knew I could build and measure with an amplifier that I had. I measured the noise spectrum from 0 to 50 kHz with an FFT-based spectrum analyzer.

Next, I calculated the real part of the impedance by dividing the mean-square value of the measured voltage noise by $4kT$. The result is the slightly jagged line in the upper figure. The smooth line is the theoretical real part of the impedance from the resistor/capacitor combination.

Then I used the measured real part and the appropriate Kramers-Kronig relation to calculate the corresponding imaginary part. Since the measurement consists of discrete points, the integral becomes a summation. The result is the imaginary part of the impedance shown as the jagged line in the lower figure. As before, the smooth line is the predicted value. I was actually surprised that it worked. Of course, if it hadn't worked, I never would have said a word about it to you. (*laughter*)

But it did work. Certainly it's a special case. You can run into all kinds of practical problems trying to implement such an approach. I was lucky enough to pick a circumstance that worked pretty well.

[Transparency I-22]

The real part of impedance also contains the essential information about power flow in the system. From the real part of the electrical impedance of a transmitting transducer, we can calculate the power input to the transducer. From the real part of the radiation impedance, we can calculate the amount of radiated power. Having these two quantities, we can calculate the transmitting efficiency. If both the noise of a system and the efficiency of a system are connected to the real part of impedance, we might guess that there's a relationship between system noise and system efficiency. In fact, there is.

For a linear, reversible transducer, reciprocity applies. If we operate the transducer in one direction (say, as a transmitter), we can calculate the ratio of the "potential" produced by some input "flow." If we operate the same transducer in the other direction (as a receiver), we can also calculate the ratio of the potential produced to the input flow. In the first case, potential is the pressure at some point in the field and the flow is the input electrical current; in the second case, the potential is the open-circuit voltage produced and the flow is the acoustic volume velocity of a hypothetical source located at the same field point. By reciprocity, these ratios are identical. By invoking this relationship for two situations, the transducer plus the medium and the medium alone, we can connect the receiving response (open-circuit voltage divided by pressure at the face of the transducer) to the transmitting response (volume velocity at the transducer face to input current). So, the first thing we find is a relationship between the transmitting response and the receiving response.

[Transparency I-23]

The pressure noise associated with the radiation impedance can be found from the Nyquist relationship involving the real part of the radiation impedance. We can convert this to the noise-equivalent voltage at the electrical terminals by the receive sensitivity, M. The total thermal noise at the electrical terminals is given by the real part of the electrical impedance. Consequently, the ratio of the noise associated with the radiation load to the total noise is simply related to the ratio of the real parts of the radiation impedance and the electrical impedance.

[Transparency I-24]

The radiated acoustic power is directly related to the real part of the radiation impedance and the total input power (as a transmitter) is directly related to the real part of the electrical impedance. We can then write the transmitting efficiency. By the reciprocity relationship, the transmitting efficiency is identical to the ratio of the noise power associated with radiation to the total noise power as a receiver. This is just one more example of the intimate association between system noise and other measures of system performance.

This is a good time for a break. Please come back in about ten minutes. While you're leaving, please pick up one of the cards with the resistor taped to it. We'll talk about that later.

II. SHOT NOISE

[Transparency II-1]

I spent the last hour talking about thermal noise. During that discussion, we didn't need to know anything about the details of the noise mechanism. Only that any form of dissipation is associated with some fluctuation. Now I'd like to talk about a mechanism for noise generation that *is* dependent on details of the process. In particular, I'd like to talk about the noise associated with the flow of discrete particles or carriers that act independently. We might consider flow of electrons or flow of photons.

[Transparency II-2]

If you believe what I've said so far, you should believe that there is a measurable voltage across the resistor in this picture. In fact, we can measure that voltage with reasonable accuracy – it represents the thermal or Johnson noise.

But let's force a flow of electrons through the resistor. Let's connect it to a current source so that a flow of discrete particles, electrons, is produced. How does the noise change? I won't answer the question right away but you'll see later that it is a very interesting question.

[Transparency II-3]

To set the stage I will give you one version of the derivation of the shot-noise expression. If we examine an electrical current closely enough, we could, in principle, observe a sequence of impulses. These impulses represent the arrival of electrons at our detector. The impulses occur at random times and they all have the same strength. So we can represent the current as a sum of delta functions.

Let's expand that sum of impulses as a Fourier series. We will pick some time interval, T, that encompasses many of these events, then expand the function as an infinite series of sines and cosines. In the cosine-sine form, we have the coefficients, a_k and b_k, and we sum over as many terms as we need for the series to adequately represent the real flow. The coefficients are obtained by integration but the integrations are simple. We're integrating over delta functions and it doesn't get much easier.

The expression for a_k is a sum of cosines sampled at the times of the impulses. The same thing is true for b_k except that the sum is of sine functions.

If the sample we take is T seconds long, then the Fourier-series coefficient (actually, the set of both a_k and b_k) represents an frequency interval of *1/T*. The mean-square value will tell us the fluctuation power in that frequency interval. After squaring, we have a cosine squared, a sine squared, and a cosine-times-sine. Under the time average, the cosine-times-sine terms go to zero and the cosine-squared and sine-squared functions go to one-half. The mean-square value is simply one half the sum of the coefficients squared.

[Transparency II-4]

Both a_k and b_k are expressed as series. To expand their squares, we write each as the product of two series. It is usually useful to separate such a product into the terms in which both indices are equal and the terms involving unequal indices. Frequently, the behavior of these two groups of terms is markedly different.

When the indices are equal, each term is cosine-squared plus sine-squared, which is one. When the indices are not equal, the result is not so clear. However, if (*and only if*) the events (the impulses) are statistically independent, these cross terms sum to zero.

I can do the sum of one from one to N – it is one of the series that I have memorized. (*laughter*) The result is simple: the mean-square value is two times the charge squared times N divided by T^2, *but only if the impulses are statistically independent*. What does statistical independence mean? It means that the occurrence of any impulse has no connection with the occurrence of any other; there's no way I can predict when the next one will occur based on any amount of knowledge of past events; or, prior events have no influence on future events.

Now, the mean value of the current is the charge times the number of events divided by the time. Also, the interval of frequency is *1/T*. Consequently, the mean-square current is 2 times q times the average (DC) current times the bandwidth. That is the classic form of the shot noise expression. There are other ways of deriving this result but this is good enough for us. (The derivation is quicker if we know something about Poisson distributions, for example.) The result applies to processes consisting of events that are impulse-like *and independent*. There are many cases in which noise processes are highly *dependent* and we will take a look at some of those processes shortly and how they differ from the shot-noise formalism.

(We can modify the derivation to remove the requirement for impulse-like events by putting in the impulse response associated with the event but it makes the derivation too difficult and does not add any fundamental insight.)

[Transparency II-5]

Let's consider a simple problem: molecules of air hitting a circular disk. There ought to be a fluctuating pressure on the surface of the disk. The molecular collisions are impulse-like so we should be able to examine the fluctuations with the shot-noise treatment.

The average force produced by the sequence of collisions is the total momentum change of a single molecule divided by the average time between collisions. That is equivalent to the molecular flux (molecules per second moving toward the disk) times the momentum change per collision.

Notice that I have drawn the molecular collision as a specular reflection. We know that the molecules do not collide with solid surfaces in the same way that light reflects from a mirror. The angle of departure of the molecule has little to do with the angle of incidence; however, I can get away with the model of specular reflection if I assume that the velocity distribution has reached equilibrium. A particular incoming molecule will not, in all likelihood, rebound specularly but I will be able to find some *other* molecule that comes off at that angle and I can pair incoming and outgoing molecules up to match angles as long as I am not changing the overall velocity distribution.

Now let's give the disk a little bit of motion and write the equation for the force on the disk. What is the molecular flux? Molecular flux is the number density of molecules divided by two – because half of them are going toward the disk and the other half are going away – times the area of that disk times the x component of the average molecular velocity minus the speed at which the disk is moving.

The momentum change is two times the mass of the molecule times the relative velocity between the molecule and the disk. Expanding this expression gives three terms. The first term is the number density times the mass times the average velocity squared times A. From kinetic theory, this is just the static pressure on the disk.

The second term is a factor times the velocity of the disk (x-dot). The third term is a factor times the velocity of the disk squared. We are going to say that the disk is not moving very fast compared to the molecular speeds. This is a very good assumption; it is very difficult to move the disk at anywhere near the speed of molecular motion. The consequence of this assumption is that the third term is negligible.

Because the second term is a factor times x-dot, the factor must be an equivalent mechanical resistance. So the mechanical resistance is two times the number density of molecules times the mass of a molecule times the average value of the x-component of velocity times the area of the disk. Now that we know the mechanical resistance, we can calculate the associated fluctuations.

[Transparency I-6]

In thermal equilibrium, the mean-square fluctuation force is simply $4kT$ times the mechanical resistance. To get the mean-square pressure, divide by the area squared. From kinetic theory, the static pressure is the number density times kT and the average molecular velocity is twice the average value of the x-component. Therefore, the fluctuation pressure is two times the average momentum change times the static pressure over area times df. This looks just like a shot-noise expression, doesn't it?

[Transparency I-7]

The nice thing about shot-noise expressions is that they all have basically the same form if you are careful how you write them. They all have the form that the mean-square fluctuation in flux density is two times the quantity of the "stuff" that the "particle" carries, whether it is momentum or charge or energy, times the average flux density times the bandwidth divided by the area through which the particles flow. For shot noise in electrical current, the flux density is the current density; the carriers are electrons with "quantity" q. Here is the shot-noise expression for an electrical current written in terms of flux densities.

It's hardly ever written this way, but it fits the model form. If we multiply through by A^2, we get the normal form for shot-noise current. For light, it's photon flux density. The photon "carries" energy, hf. The optical intensity, I, is the flux density.

And the third example is the one we just found. Pressure is momentum flux density. Consider this expression more closely. This equation says that the mean-square fluctuations in pressure depend linearly on the static pressure. In other words, the noise power in the pressure fluctuations is directly proportional to the static pressure. This is a very important conclusion *because this linear dependency is NOT observed in measurements at normal pressures*. Apparently, something is wrong.

[Transparency II-8]

Well, let's try the derivation another way. For very small disk speed, we can use the Stokes' flow solution to calculate the mechanical resistance experienced by the disk. In that case, the mechanical resistance is 16 times the viscosity of the fluid times the radius of the disk. Once we have an expression for the resistance, we can calculate the pressure-fluctuation noise power using Nyquist's relation. However, by doing this, we find that the mean-square pressure fluctuations are *not* proportional to static pressure. (Actually, viscosity is a function of pressure for real gases but the dependence is very weak.) So we have a problem – which solution is correct?

[Transparency II-9]

Remember I said that for the shot noise derivation to be accurate, the intermolecular collisions must be independent. That is, in fact, the source of the problem. Take a look at the ratio of the two expressions we obtained for the mean-square pressure. The ratio is a constant (nearly equal to one) times the radius of the disk divided by the molecular mean free path. So there is a very interesting transition when the mean free path – the average distance between intermolecular collisions – becomes equal to the radius of the disk.

When the mean free path is much smaller than the radius of the disk, a molecule cannot simply come in from far away, collide with the disk, then travel away again. Instead, it is constantly interacting (colliding) with other molecules. Intermolecular collisions are more influential on the velocity distribution of the molecules than are collisions with the disk. The collisions are not independent events. The motion of the disk modifies the velocity distribution of the rebounding molecules but those rebounding molecules almost immediately transfer those changes to the incoming molecules. The incoming molecules are strongly influenced by the reflecting molecules. The fundamental premise of the shot-noise development is wrong.

However, when the mean free path is very large compared to the disk radius, then the derivation is fine. The molecules do not interact very often with each other compared to their probability of hitting the disk and an incoming molecule has very little chance of encountering a rebounding molecule before striking the disk.

If I plot the fluctuation pressure as a function of static pressure, then at high pressure (meaning pressures for which the mean free path is small compared to the disk radius) there is little dependence on static pressure. For low pressures, however, there is the expected linear dependence on pressure. This latter regime is the free-molecular-flow regime; the former is the normal continuum regime.

Interestingly enough, from measurements made for flow through tubes, at atmospheric pressure there is a 10% departure from the continuum assumption for a tube diameter of 6 micrometers. The mean free path in air at atmospheric pressure is a bit less than 0.1 micrometers and you'd think that 6 micrometers is much larger than that. But we start to see the effects of independence well before we reach the fully developed free-molecular-flow regime. This "pre-free-molecular flow" region is called "slip flow," because the flow behaves as if it is continuum flow but with slipping at the boundary – the normal assumption in continuum viscous flow is that the tangential flow speed is always zero right at the boundary.

This has consequences for microfabricated sensors because the spacing between moving structures can often be just a few micrometers. Consequently, you cannot be too glib about the continuum hypothesis in these devices; you have to understand carefully how the molecules are behaving.

You certainly have to be careful to determine whether discrete events are independent; it's tempting to assign any process involving discrete carriers a shot-noise component and not all of those processes are accompanied by shot noise.

MR. APOSTOLOU: Are there sensors that can be evacuated to reduce the noise associated with molecular impact?

DR. GABRIELSON: Definitely. A pressure sensor must be in contact with the fluid in order to function and we can't avoid the collision mechanism; however, an accelerometer can be packaged in a vacuum. This only removes the fluctuations associated with molecular collision, though. We still need to consider other loss mechanisms and we also need to consider the impact of electronics. I'll say more about electronics in the third hour.

[Transparency II-10]

Now, let's consider another flow of discrete objects: electrical current. First, we'll look at electrical conduction in metals. Metals are characterized by a conduction band, which is populated with plenty of carriers (electrons). Conduction bands in metals are the urban areas of electrical conduction. The population density is high.

Because population density is so high and because no two electrons can occupy the same state (the exclusion principle), their motion is strongly correlated. They cannot just jump randomly into any nearby state; for the most part, the nearby states are full. They can only move if another one moves to vacate a state (which, of course, means that many other carriers must be involved to provide suitable vacancies). The carriers can move together, but there is a substantial coordination in the movement.

(The classical way of making this argument is to point out that each electron has an electric field and, if they are close together, the interaction of the electric fields produces a strong correlation in the motion from electron to electron.)

As a consequence, the carriers in a metal are highly dependent: we do not observe shot noise in metallic conduction. The fundamental noise mechanism is the thermal Johnson noise.

[Transparency II-11]

Semiconductors are a different story. If we consider an intrinsic semiconductor, all of the carriers in the conduction band got into the conduction band by being thermally excited from lower energy levels. Unless the semiconductor is very heavily doped, the density of conduction carriers is very low compared to a metal. There are plenty of unoccupied neighboring states and the carriers can hop from state to state without much influence on other carriers. Consequently, for conduction in a semiconductor, we do observe shot noise.

[Transparency II-12]

I would like to explain this same phenomenon in a different way. Remember that, when I was discussing the noise associated with radiation, I drew a plot in *k*-space for the acoustic modes in a box. We can draw the same type of plot for the states occupied by the electrons.

Because of the exclusion principle, we build the diagram up by adding electrons to the k-space grid with one electron per grid point (two, if you'd like to account for both spin states). We fill up the states from the low-momentum states near the center outward as we add more and more electrons.

The occupation diagram in *k*-space for a metal looks like a solid disk centered on the origin. The occupation extends out to rather large values. For copper, the outer edge is at about 7 electron volts whereas the average thermal energy (*kT*) at room temperature is about 0.25

electron volts. At zero temperature, the edge of the disk is very sharply defined. There would be no fluctuations to blur the edge. If it looks fuzzy to you, then you need to visit the eye doctor.

What happens if we raise the temperature a bit? The states in the interior of the disk are fully occupied. Even at room temperature, the available thermal energy is only 0.025 electron volts so an electron occupying a state in the interior doesn't have enough energy to jump to the nearest unoccupied state at the disk's edge.

So, at elevated temperature, the occupation picture looks like a slightly out-of-focus version of the same disk. The disk is fuzzy around the outside. It is still solid on the inside – the occupation is still complete on the inside – but now the outer edge is fuzzy.

[Transparency II-13]

How do the conduction carriers in a semiconductor appear in *k*-space? An intrinsic semiconductor at zero kelvin has no carriers so *k*-space is unpopulated. As the temperature increases, some electrons (or holes) are sufficiently excited thermally to reach the conduction band. There is a small, blurry region around the origin; the population density is small. On the overhead, it looks like a quantum dust bunny. As in our previous analysis, nearby states are easily accessible so the carriers can act as if they were independent.

If we apply an electric field, then the distribution translates one way or another, depending on the direction of the electronic field. The direction of translation corresponds to the flow of current and the fluctuations constitute a feature as significant as the shift. This is characteristic of shot noise. The magnitude of the fluctuations is proportional to the steady flow.

If we had applied a field to the metal, the occupation disk would also have shifted. However, because the radius of the disk is extremely large compared to the thermal excitations available, the pattern of fluctuation (the fuzzy edge of the disk) doesn't change noticeably. This is characteristic of Johnson noise – the fluctuations do not depend on the bulk flow.

[Transparency II-14]

This is not just interesting theory. The difference between the shot-noise prediction and the Johnson-noise prediction for a metallic conductor is so large that it's relatively easy to demonstrate with a simple measurement. This is a useful (although sometimes impractical) strategy for physical investigation: you get some insight, translate that insight into a measurement, and confirm (or refute) the insight. If the measurement fails to correspond to your insight, then you don't tell anyone about it; if it confirms your insight, then you make a big deal out of it. (*laughter*) Actually, as Feynmann (I think) suggested, you should commit to publishing the results regardless of the outcome. If you do careful experimental work, the results are valuable either if they confirm what you think should happen or if they run counter to what you think. Unfortunately, many people do just what I suggested in jest. You're probably going to be wrong more often than you'll be right. As long as you learn the lesson nature is trying to teach you've made a valuable contribution either way.

Enough philosophy. I construct this simple circuit. It's little more than a standard, non-inverting amplifier with a gain of about 100 except that the usual input resistor on the positive input has been replaced by a bridge of four, one-megohm resistors with battery and a switch. Regardless of whether the switch is open or closed, the equivalent resistance of the bridge is just one megohm. No matter what I connect between the two points where the battery is drawn, the amplifier sees the same input resistance so its operating point doesn't change. The only input to the amplifier is noise produced by the resistor network. The amplifier is an inexpensive but fairly low-noise operational amplifier. A noisy amplifier chip would mask the noise produced by the resistor network.

The bridge arrangement is convenient for this measurement. If the bridge is well-balanced (nearly equal resistance in each leg), then even a large voltage applied at the mid-point connections does not produce any voltage at the input to the amplifier. Also, noise produced by the battery flows equally in both legs of the bridge, so that battery noise is cancelled to a large degree. What we are interested in measuring is noise produced in the resistors – either the basic thermal noise or additional noise produced by current flow. The bridge arrangement allows us to measure this without contamination from the source of voltage.

If you picked up one of the little cards with the attached resistor that I passed out, you have the official Penn State precision noise source. It is really quite good. It is a one megohm resistor with a 1% tolerance. If you know what the temperature is, you can predict the thermal noise level produced by the resistor. If you're careful, you can also measure that noise level. The circuit I show is suitable for such a measurement. You could even connect it directly to a good Stanford lock-in amplifier and measure the noise. Most spectrum analyzers are not capable of measuring that noise level directly but, using a simple, low-noise amplifier, you can. We obtain the same result from the bridge. The noise is flat over a reasonable frequency range (it is "white"). If we close the switch to force current through the resistor network, we can again measure the noise.

[Transparency II-15]

The curve labeled *B* is the noise measurement of the resistor bridge with no current flowing through it. Shown to the right is the level predicted by the Nyquist expression for thermal noise.

If I close the switch using a 9V battery, the noise changes somewhat at low frequency. This measurement is given by the curve labeled *A*. If I calculate the noise expected according to the shot-noise formula, the level is an order of magnitude higher than the measurement. This is a convincing demonstration that shot noise is not an operative process in for conduction in a metal-film resistor, and it is something you can do at yourself without sophisticated equipment.

In our simple theory, the noise with current flowing should be virtually the same as without. However, at low frequency, there is more noise when current is flowing. There are two reasons for this. The extra wire associated with the connection to the battery picks up more interference of 60 Hz and its multiples than when the battery is not used. More important, there is another noise mechanism that we have not discussed. Frequently, when power is being supplied to a system, noise is produced that has a spectral distribution in power that varies as the inverse of frequency. This noise component is not straightforward to model but it is often observed. For certain types of resistors (carbon-composition, especially), the so-called *1/f* noise can be substantial. For the metal-film resistors used in this measurement, it's rather small but still measurable.

[Transparency II-16]

Next, let's examine a case of electrical conduction in which the carriers *are* independent. We'll use the model of an electrical diode made in the usual way with a *PN* semiconductor junction. We'll apply a small voltage to the diode and examine the behavior of the current.

The *PN* junction forms a potential barrier. Regardless of the sign of the voltage, the current can be represented as the sum of a forward current and a reverse current. We can't measure the two currents separately but we can describe them separately. The reason that we can postulate two currents is because of fluctuations – charges will "flow" against the potential on occasion because of their thermal energy.

The forward and reverse currents are related by the most important result from statistical mechanics – the Boltzmann distribution. The probability of a carrier traveling "uphill" with respect to the applied voltage is related to the probability of a carrier traveling "downhill" by an exponential. The exponent is the ratio of the potential energy difference from one side of the junction to the other to the unit of thermal energy, kT.

The current that we would measure with a DC current meter is I, the difference between the forward and reverse currents. Using the Boltzmann relation, I can write the measurable current in terms of the forward current. Ohm's law – the voltage difference divided by the effective resistance – also gives the measurable current.

The noise associated with the forward current is not correlated with the noise associated with the reverse current. The passage of charge across the *PN*-junction potential barrier is a flow of independent carriers. So the total noise power (expressed as the mean square fluctuation in current) is the sum of the squares of the current fluctuations from the forward and reverse current. The total current noise is then the sum of two shot-noise expressions. We can eliminate the forward current by substituting ΔV over R and the exponential relation above. The total noise is this rather unwieldy expression involving the charge, q, the voltage drop, the resistance, and a combination of exponentials. Of course, you can program it and graph it, but it is often useful to examine the limits and that is what we will do.

[Transparency II-17]

For a $q\Delta V$ much smaller than kT, that is when the voltage across the diode is much smaller than about 25 mV (at room temperature), the expression goes to a Johnson-noise limit. (Remember that, for current noise, the Johnson-noise expression has the resistance in the denominator.)

In the "large" voltage limit – when the voltage across the diode is much larger than 25 mV – the limit is the shot-noise expression. So it all makes sense. When I do not apply any external voltage, the noise is simply Johnson noise, and that's what should happen. When I force a current of independent carriers through the junction, the noise is shot noise. There's a nice transition between the Johnson-noise expression and the shot-noise expression and that's described by the more complicated form on the previous slide.

I've included a figure from an interesting paper. The paper was published in 1997 and the subject was the search for the fractional quantum Hall effect. It doesn't matter to us just what the fractional quantum Hall effect is. The important aspect is the prediction that, in a certain special circumstance, the electrical charge carriers will appear to have a charge of $q/3$ instead of q. How did these people test this hypothesis? By measuring noise. Very nice.

The plot shows the spectral density of the noise that they measured as a function of the DC current. The low-current limit gives them the Johnson-noise value and this value serves as a calibration of their measurement independent of the fractional-charge effect. As the current is increased, the noise power makes a transition from independence of current (Johnson-noise regime) to linear dependence on current (shot-noise regime). But the slope of the shot-noise section of the curve corresponds to $q/3$, not q. Fascinating: a fundamental physical measurement with noise.

[Transparency II-18]

This last experiment regarding the fractional quantum Hall effect is very interesting but so are some more common processes. Let's consider the noise in transistors. This subject is not as arcane as the quantum Hall effect but it's still rich in physics and, normally, more practical. We stray a bit from sensors, acoustics, and vibrations but if you make a sensor, you usually process the signal by means of electronics.

I'm going to discuss noise process in the bipolar junction transistor. There are several important varieties of transistors but the noise analysis for the bipolar transistor is straightforward. The transistor has a base, collector, and emitter. There is a relatively large current that flows from collector to emitter (or vice versa) and a relatively small current that flows into (or out of) the base. The small base current controls the large collector-emitter current and this is the basis of the device as an amplifier. The collector current is roughly proportional to the base current through the parameter, β. Since β is much larger than one (for transistors used to amplify small signals), the collector and emitter currents are very nearly equal.

Various models are used to represent the behavior of transistors. On the right, I've drawn a simple version of the transconductance model. This model represents the transistor as a voltage-dependent current source. The collector current is equal to a constant – the transconductance, g_m – times the base-emitter voltage. The unit of transconductance is inverse resistance.

Because the transistor junctions look just like diodes, we can use the expression we derived for the *PN* junction to write the collector current. If I calculate the transconductance by differentiating the collector current with respect to the base-emitter voltage, I obtain the quantity, qI_c over kT. This assumes that the base-emitter voltage is much larger than 25 millivolts, but this is usually a good assumption. Because the current produced (I_c) is in phase with the control voltage (V_{be}), the reciprocal of transconductance really does represent a true resistance. This is important because resistances or mechanisms that behave just like resistances generate noise.

[Transparency II-19]

Now that we have a model for the transistor, let's calculate the noise. We'll calculate the noise from the point of view of the input (the base) to the transistor – this is the conventional way to reference the noise in a device. There is a shot-noise component associated with the base current. Because the base-to-emitter is a reverse-biased junction, the carriers are independent and we get full shot noise: $2qI_b$ or $2qI_c$ divided β.

Since the transconductance represents the reciprocal of a resistance-like phenomenon, there is a voltage noise associated with it. Because the voltage is in phase with the current, it does not really matter whether you can identify a physical part having some measurable resistance. What's important is that the phenomenon is dissipative. The voltage noise power is $4kT$ times the reciprocal of the transconductance and that can be rewritten in terms of q and the collector current.

An interesting measure of the noise power of a device is the product of i_n and e_n divided by the quantity, $4kT$. This number is often at least roughly constant for a particular device and, if the number is significantly less than one, the device can be characterized as "low noise." For the bipolar transistor, the result is one over the square root of 2β. Although it is not obvious from our discussion, β is not a strong function of the operating point of these devices – it is roughly constant. If the collector current changes by several orders of magnitude, β might change by a factor of two. For a given transistor geometry and material, you are stuck with this number. If you increase the collector current, you can reduce the voltage noise but the current noise goes up; if you decrease the collector current, the current noise decreases but the voltage noise increases. This is a common tradeoff.

The next quantity is more common in the literature. People often divide the noise voltage by the noise current – they call it "noise resistance" – to get a function that does depend strongly on the operating point. It is inversely proportional to the collector current. The "noise resistance" is often taken to be the optimum source resistance for best noise performance in the matching of a sensor (the "source") to a transistor but the situation is actually more complicated than that. We'll talk more about this later.

In a real transistor, of course, we also have to worry about physical resistances in the transistor structure. If there is a real physical resistance, we must calculate the noise associated with that also.

[Transparency II-20]

What are the consequences for transistor design then? Let's try for a voltage noise of 1 nanovolt per root hertz. That happens to be a pretty good specification for noise voltage; devices of this sort are commercially available. From the fundamental transistor relations, I know that the collector current must be 0.4 milliamps to get that noise value. Notice, from the fundamental relations, that β doesn't enter the calculation for voltage noise; the voltage noise level is a fundamental quantity. I would need to know β to find the noise current, but I don't need to know β to find the noise voltage.

In an operational amplifier, you would need two transistors for the differential front end, so the total current consumption just for these two transistors would be almost one milliamp. This is quite consistent with the current consumption of op amps with this noise voltage. Also, the effective emitter resistance is 60 ohms, so it's not too difficult to keep the physical resistances below this value. (Remember that we have to account for Johnson noise from any physical resistances also.)

Now let's try for an equivalent input voltage noise of 0.1 nanovolts per root hertz. As far as I know such transistors are not commercially available but why not try to build one – there'd be a market for it. What is the required collector current now? Forty milliamps for each transistor or almost 0.1 amps for an operational amplifier front end. Wow.

Commercial power transistors can handle currents that high but at these current levels we need to consider thermal dissipation in the chip. Also, the effective emitter resistance is six tenths of an ohm, so we also need to be very careful to keep physical resistances in the base-emitter path low. This requires a massively parallel base architecture – the base region would need to be very wide to keep the resistance low. In practice, this is accomplished by placing many structures in parallel, which takes up considerable space on the chip. I don't think it is out of the question to build a transistor with this noise level but it would be very large and would probably need a heat sink. Wouldn't that be interesting: a low-noise amplifier with a heat sink on the input stage?

[Transparency II-21]

But modern transistors are so small does a "large" one really take up much room? Some of the integrated circuit manufacturers publish their die layouts so that people who want to use the bare die instead of the packaged integrated circuit can figure out where the connections must be made. One example is the Analog Devices AD743, a good representative of the class of low-noise amplifiers.

Here is the die layout. On the left are the input transistors. It actually looks like there are four grid-like devices but it's really only two. The two are interleaved so that processing differences in production and thermal gradients in operation average out across the two input transistors. (This is called "common centroid" design.)

These two input transistors occupy an area of one millimeter by one millimeter. That's a density of about two transistors per square millimeter. That's the area required for thermal dissipation and for low base resistance. In fact, this chip when packaged comes in an 8-pin standard package (the "DIP"); however, it is only available in the 16-pin small-outline package because the die won't fit into the 8-pin small outline. This is a substantial design constraint that many people overlook. We are so used to the incredibly high densities of transistors used in microcomputer chips that we're accustomed to thinking that the real estate required for a single transistor is entirely negligible. Pentium chips have thousands of transistors per square millimeter.

It pays to understand the size requirements of low-noise design. The technology known as MEMS (Micro-Electro-Mechanical Systems) is hot right now. Many people are proposing extremely small mechanical structures, sometimes for sensing, sometimes for actuating. They are so fascinated with the new freedom in creating very small mechanical structures that they don't worry about the support electronics. But, if they are designing a sensitive sensor, it may turn out that the area required for the support electronics is far larger than for the mechanical structure. The low-noise electronics may completely dominate the size of the device and the great advantage of building small is lost. You always have to be careful about glib statements concerning new technologies.

That brings us to the end of the second hour. This is a good place to stop and take a break.

III. SPECIAL TOPICS

[Transparency III-1]

At the beginning of this lecture, we discussed the most fundamental of noise processes – the noise generation that results from random thermal agitation of molecules. This source of noise is relevant to almost any problem we might consider. Then we considered shot noise, which still has fundamental roots, but is not applicable as broadly as thermal noise. Just before the break, we extended these ideas to transistors and now we are going to discuss a few more extensions of these basic ideas. The underlying theme is that these fundamental physical processes show up in interesting ways and in interesting structures.

[Transparency III-2]

From a practical point of view, one of the most important subjects is the interaction between electronics and a sensor. We usually connect our physical sensors to some kind of interface and let's examine the classical method for analyzing this problem.

The standard model for an amplifier is constructed as follows. We replace the real amplifier with a noiseless gain stage and two equivalent noise sources – an equivalent voltage noise source and an equivalent current noise source. There are, then, two basic components to the noise associated with the electronics. One component is simply the effective voltage noise of the amplifier; the other is produced by the amplifier current noise flowing into the sensor impedance thereby generating another voltage-equivalent component. Then, of course, there is the basic Nyquist noise associated with the sensor. I won't prove that this is adequate but I will give you a plausibility argument later.

It is important to understand the implications of this model when designing systems for maximum performance. For example, if the sensor is predominantly capacitive, the equivalent current noise is critical. At low frequency, the impedance of the capacitor is high and the current-noise component can dominate; at high frequency, the impedance is low and the voltage-noise component can dominate. The notion of one "best" amplifier usually leads to suboptimal designs, something we want to avoid.

Once we've minimized each component's contribution to the total system noise as far as possible and the only remaining strategy is to reduce Boltzmann's constant, we've completed a successful design. (*laughter*) Of course, there are some circumstances in which we can reduce the temperature for some reduction in noise but, if the noise is fundamental thermal noise, going to liquid-nitrogen temperatures only buys a factor of four in temperature or a factor of two in noise amplitude.

[Transparency III-3]

At Penn State we teach a transducers lab course; one of the labs involves measuring the noise characteristics of resistors and amplifiers. Here is a picture of one of the breadboards with an integrated circuit, a couple of gain resistors, and a place to plug in other resistors. Alongside is the schematic diagram showing a simple non-inverting amplifier with a gain of 101.

By changing the resistance value at R_s, we change the relative proportions of the Johnson noise, the amplifier voltage noise, and the noise voltage produced by the noise current flowing into the resistor. Also, we perform the same set of measurements with different kinds of amplifier chips.

[Transparency III-4]

Here is an example of such a set of measurements performed with two different amplifiers. One curve is for the AD743 that I showed you before. When we put various resistors on the positive input, we discover that for low values of resistors the output noise does not depend on the value of the resistor. The noise stays at some level and that level is, in fact, the voltage noise of the amplifier. (The values shown on the plot are the values referred to the amplifier input. Multiply by 101 if you want to find the actual output noise.)

For another range of resistor values, we discover that the measured noise is very close to the Johnson noise of the source resistance. Then, for very large values of resistance, we discover that the curve again bends away from the line that represents the resistors' Johnson noise.

In low-noise amplifiers there are typically three distinct domains of operation depending on the value of source resistance. We see one region in which the noise does not depend on resistance, one region where the noise increases as the square root of resistance, and one region where the noise is linear with resistance.

If we examine another chip – the LM4250, a low-power chip – and perform the same sequence of measurements, we discover a similar behavior. The most obvious difference is that the total noise is always higher than that of the AD743. Also, there is only a very small region in which the noise curve approaches the resistor Johnson-noise curve. The LM4250 is not marketed as a low-noise amplifier; it is marketed as a low-power amplifier.

Let's calculate some of the quantities that we discussed earlier in relation to the transistor. The product of e_n and i_n divided by $4kT$ is 0.42 for the LM4250. For the AD743 it is 0.001. The smaller this number is, the wider the range of resistance for which the amplifier's noise is below the intrinsic resistor noise. The ratio of e_n and i_n is 240 $k\Omega$ for the LM4250 and 420 $k\Omega$ for the AD743. Notice that, at 240 $k\Omega$, the LM4250 curve reaches its point of closest approach to the intrinsic resistor-noise curve.

If the normalized $e_n i_n$ product is greater than about 0.1 and the sensor's impedance is predominately resistive, the e_n/i_n ratio can be used to select the amplifier by matching it to the sensor resistance. Before we had such a variety of very-low-noise chips, this used to be the recommended way of selecting amplifiers. It is not a good design procedure these days but, unfortunately, it is still described as a design procedure in recent literature.

With the AD743, there is a large region where the total noise is dominated by the intrinsic resistance noise. The smaller the normalized $e_n i_n$ product is, the larger that region is. In fact, if you take the e_n/i_n ratio and multiply by the normalized $e_n i_n$ product, you get 420 ohms, which is the resistance at which, on the low end, the total noise curves starts to depart significantly from the Johnson-noise curve. If you take the e_n/i_n ratio and divide by the normalized $e_n i_n$ product, you get 420 megohms, which is the resistance at which, on the high end, the total noise curve starts to depart. So, we can easily predict the range of resistance over which the amplifier's contribution to the total noise is negligible compared to the intrinsic contribution (for a resistive source). The width of that region for the AD743 is very large whereas the width of the region for the LM4250 is very small. If we were using a sensor with a 240 $k\Omega$ resistance and we were designing by the rule of matching the e_n/i_n ratio, we'd pick the LM4250 over the AD743. And we would be making a very poor choice! For a source resistance from 10 $k\Omega$ to 100 $M\Omega$, the AD743 would be a reasonable choice even though those limits are far from the value given by the e_n/i_n ratio.

I want to emphasize that this is not a totally theoretical process. The measurement is definitely practical. With some care in constructing the circuit and shielding, it is suitable even for an undergraduate lab exercise. In fact, in our lab we used to do it using those white plastic perforated plug-in boards, which are really not suitable for many precision measurements. There would be a great deal more understanding of the contribution of electronics to system noise if more people performed this experiment.

[Transparency III-5]

I sorted through some manufacturers' specification sheets and I recorded values for equivalent voltage noise and equivalent current noise (at 1000 Hz) of a number of operational-amplifier chips and discrete transistors. Then I made a scatter plot of those values. From this plot, you can see that it is hard to find a device with equivalent voltage noise below a nanovolt per root hertz. Those devices that are close to one nanovolt per root hertz are based on bipolar transistors.

Also, it is hard to find a device anywhere in the region below 2 nanovolts per root hertz *and* below a few hundred femtoamps per root hertz. The bipolar devices cluster in one region with low voltage noise and relatively high current noise. The junction field-effect devices cluster in another region: they have lower noise currents but somewhat higher noise voltage.

MOS and micropower devices do not have very low voltage noise but the MOS devices can reach reasonable levels for current noise. A plot like this provides some direction for making intelligent choices. If power consumption is really important, I'm going to have a hard time reaching a nanovolt per root hertz. If I can back off on the voltage-noise requirement, then I can work with the very-low power consumption devices.

It's important to understand that this plot is constructed for a single frequency – in this case, 1000 Hz. At low frequency, noise with a 1/f power spectrum enters the picture. A device with one nanovolt per root hertz at 1000 Hz may show much higher levels at one hertz. Both the current and voltage noise may show 1/f behavior at low frequency. (These effects are usually most pronounced in the MOS devices.)

Another important factor is temperature. From the basic equations for the bipolar transistor, there is a fundamental temperature dependence in the noise. For junction field effect transistors, temperature effects can be even more extreme. The noise current in a JFET device is dependent, through a shot-noise relation, on the steady leakage current into the gate (the analog to the base of a bipolar transistor). This leakage current is exponentially dependent on temperature. Consequently, for increasing temperature, this leakage current can become quite large with a corresponding increase in current noise. A specification for current noise that is attractive at 20 degrees C may translate to unacceptably high current noise at 50 degrees C.

[Transparency III-6]

I can take the same data and plot them on other axes. If I use the factor of $e_n\, i_n\, /\, 4kT$ for the ordinate and the ratio e_n/i_n for the abscissa, the plot emphasizes other features or limitations of the devices. The ordinate is something akin to a figure of merit for the amplifier; and the abscissa is the value corresponding to the center of the range of "optimum" source resistance. You can see immediately that the junction field-effect transistors are the clear winners in terms of our "figure of merit." They can be used effectively over fairly large ranges of source impedance.

Now it is useful to see the variations in noise performance of real devices, but, more importantly, we have learned something about why there are limits to these devices; more fundamental limits than manufacturing process limits.

[Transparency III-7]

Now that we have some practical tools for noise analysis, let's consider a system with a bit more complexity. About a year ago, we were interested in developing a microfabricated accelerometer. We did not want to develop a fabrication process, so I located a manufacturer that made a packaged accelerometer. The noise floor of the commercial accelerometer was too high for our application, but it appeared that the integrated electronics was the limiting factor rather than intrinsic limits in the mechanical structure. The manufacturer sold me a few of the mechanical structures and I built an accelerometer based on those structures.

Each of the structures was a differential capacitor. Each had a central moving mass between two fixed plates so that when the mass moved the capacitance on one side increased and the capacitance on the other side decreased. For my device, I used two of these structures so that I could assemble a full bridge – two capacitances increased and two decreased with motion of the center masses.

I drove the bridge with a 40 kHz sine wave generated by a function-generator chip. The upper side of the bridge was driven with the sine wave and the lower side was driven with an inverted replica of the same sine wave. Consequently, if there were no displacement of the center masses, there would be no output. With some displacement of the center masses, there would be an output at 40 kHz. The amplitude of the 40 kHz output is proportional to the displacement of the center mass set.

Instead of generating a voltage that is directly proportional to the acceleration input, this configuration produces a 40 kHz signal that is amplitude modulated according to the acceleration. This technique has been around for a long time and there are some significant advantages to the procedure even though it adds some complexity. The output of the bridge goes to a differential input stage then to a high-pass filter. The high-pass filter eliminates much of the additive electromagnetic interference.

Then the signal goes to one input of an analog multiplier. A replica of the signal that is used to drive the bridge goes to the other input. After low-pass filtering the output of the multiplier, the actual acceleration signal is recovered. This process is analogous to the process in a lock-in amplifier but without the frills.

In a complicated system, it is useful to analyze each component to determine its contribution to the overall noise. Then, if one component is found to dominate, that's the component to attack in lowering the noise. Once you've gotten down to the fundamental limits you're done unless you can change the sensor itself. I show, on this slide, the noise budget for this system. I've referred everything to the output. For this system, it turns out that the limiting noise is in the multiplier (since I can't do anything about the basic sensor chip).

It's worth making a point here. Many people think that if they have analyzed the first stage of the electronics and everything is fine, they do not have to do any more than that. However, sometimes the noise is dominated by a stage further down the line as it is in this example. The commercial off-the-shelf accelerometer has a noise floor of 25 micro-g per root hertz. My measured performance was about 0.5 micro-g per root hertz, almost two orders of magnitude better.

[Transparency III-8]

Why would you consider such a complex system including a high-frequency drive and synchronous detection? The bridge would work with a DC voltage. There are many advantages. Earlier I said that for a sensor that is electrically a capacitor there is a component of noise that increases rapidly with decreasing frequency. This component results from the current noise of the amplifier flowing into the ever increasing capacitive impedance. Because it's difficult to make a small air-gap capacitor with high capacitance, this noise component can dominate easily at low frequency.

Also, much of the electromagnetic interference appears at low frequency. If important sensor signals are in the same frequency region, then it's difficult to get good performance. If, however, I translate the important signal information up in frequency, I'm able to separate the electromagnetic interference from the signals. Furthermore, the relevant noise characteristics of the first-stage amplifier are those characteristics at 40 kHz. So the problem with current noise into the capacitive impedance is not an issue nor is 1/f noise in the amplifier.

In fact, the resistance to electromagnetic interference is impressive with this technique. In the transducers lab class at Penn State, we set up a strain-gauge bridge on an aluminum cantilever. We wrap the bare output wires from the bridge around a ballast transformer for a fluorescent light a few times and *then* into the first-stage amplifier. Fluorescent ballast transformers are notorious for generating high levels of electromagnetic interference so this is a pretty severe environment. But, the high-frequency drive, filtering, and synchronous detection make the final signal very clean.

DR. GARRETT: Does the term "sadist" mean anything to you? (*laughter*)

DR. GABRIELSON: Yes, it's synonymous with "professor" at Penn State. (*laughter*)

These techniques are important to keep in mind. If you have the basic understanding of synchronous detection, you have an approach to reducing the system noise. If you understand the fundamental limits, then know how far it is practical to go in noise reduction. Building up your toolbox of techniques is a very useful exercise, one that can pay dividends when you are faced with difficult problems.

[Transparency III-9]

Now I would like to talk about noise in fiber-optic sensors. Fiber-optic sensors are a wonderful place to look if you are interested in noise, because they involve so many different kinds of noise mechanisms. If you're not aware of them, they'll frustrate your attempts to build good sensors.

For years, many people have said that fiber-optic sensors are wonderful. Optical fiber is cheap and all I need to do to increase the performance is wind more fiber. Any time you hear someone say "all I have to do is ...," you should know that there's trouble ahead. Often, it is the thing that we think is trivial that cause the most trouble when venturing onto new ground.

As I said, there is a wealth of noise mechanisms. We've already talked about thermal fluctuations in the sensor structure. We have also considered the noise associated with electronics. Optical detection involves shot noise – the photons are independent unless we raise the optical intensity so high that we are in danger of melting the optical detector. And we already have the tools to analyze shot noise. Beyond shot noise, excess noise can be problematic.

There are noise limits associated with demodulators. In interferometric sensors, we need to convert phase shifts to voltage level changes. My friend at the Naval Research Laboratory, Tony Dandridge, tells me that 1 microradian per root hertz equivalent phase noise is about as good as you can do in the laboratory. For a commercially viable system, you wouldn't plan on doing this well.

The light source that drives the interferometer contributes noise. Lasers have both amplitude noise and phase noise (and some other weird things that are fun to learn about). Thermal fluctuations in the fiber contribute noise by generating extraneous phase shifts. Coherent Rayleigh scattering (CRS) and stimulated Brillouin scattering (SBS) also corrupt signals – CRS when signals traverse fiber in both directions and SBS when the light levels are high enough to stimulate emission in the fiber. So much material, so little time. I'm only going to discuss shot noise in the photodetection process and thermal fluctuations in the fiber.

[Transparency III-10]

There are a number of ways of figuring out what thermal agitation in the fiber does to an interferometric sensor. First, what are we worried about? We are worried about phase changes – phase fluctuations – in an interferometric sensor. In a pressure sensor, we worried about pressure fluctuations. In an acceleration sensor, we worried about acceleration fluctuations. For the interferometer, we need to find out what the magnitude of phase fluctuations is. There are several ways of calculating this; I am going to show you one way.

Let's start with an analogy. Electrical power is voltage times current (the in-phase components but let's not worry about that now). What is the analog for thermal power? Heat power, Q-dot – or dQ/dT – is equal to temperature times S-dot, the time derivative of entropy.

The Q-dot is power, the potential is temperature, and the rate of change of entropy is the flow quantity. I'm not going to say the word "entropy" again. Don't be scared by it, just consider it the flow quantity in this equation.

For a cylindrical element with radius a and length L, Q-dot is equal to the thermal conductivity times the cross-sectional area times the length times the potential difference across the cylinder – the ΔT. The analog for an electrical conductor is that the electrical power is the current, i, times the potential difference, Δe. To find the flow quantity in the thermal case, we divide the power, Q-dot, by the potential, T.

For the electrical element, we could perform the equivalent operation but we also know that the flow, i, is equal to the potential divided by the resistance, R. By analogy, the flow in the thermal case must equal the potential divided by an analog to the electrical resistance. The thermal "resistance" is then L times T divided by the quantity thermal conductivity times area. This is not the usual definition of thermal resistance but it is the proper analogy to the electrical case.

Now that we know the equivalent resistance, it's child's play to find the potential (temperature) fluctuations. We simply use the Nyquist relation. Of course, talking about temperature fluctuations is dangerous. Temperature, itself, is defined in terms of equilibrium states. I'm using the concept of potential fluctuations as a means to an end, though, so bear with me.

[Transparency III-11]

If we consider a piece of optical fiber, we would find that the local temperature at any point changes randomly. Very nearby, the temperature fluctuations would be very similar. Further away, there would be some similarities and some differences between the patterns of fluctuation. Even further, there wouldn't be much similarity between the fluctuations. This suggests a particular scale length for the "coherence" of temperature fluctuations. What do you think the scale length is for that phenomenon – for the coherence length of temperature fluctuations? Any guesses?

DR. MIGLIORI: The thermal penetration depth.

DR. GABRIELSON: Give that man a cigar. Yes, the thermal penetration depth. It is not only an essential parameter for thermoacoustics, it is also relevant for noise analysis. For an interferometric sensor, phase change as a result of transit through the fiber is important. The phase change associated with a "patch" of temperature fluctuation is the change in wave number that results from the temperature fluctuation times the coherence length. Or, written more conventionally, the change in index of refraction times the average wave number times the correlation length.

Well, what is the change in index of refraction? It is the change in temperature times the derivative of the index of refraction with respect to temperature. The index itself is temperature-dependent but, in addition, the temperature changes the length; both effects must be included. So there are two terms, one is the fundamental change in index of refraction with temperature and the other is the change in effective index because of the change in length.

The mean-square phase fluctuation then equals the mean-square temperature fluctuation times the square of the change in index with temperature times $(k_0 l)^2$. Of course, I can substitute our previous expression for the mean-square temperature fluctuation. It is getting a little ugly but we'll be done soon.

[Transparency III-12]

What I am going to try to do is to try to extract the essence of the behavior and forget about the details in the equation. Nice when you can do it. The basic size scale, the thermal penetration depth, determines the nature of the fluctuations. The fiber appears to act as if it were a collection of patches each with dimension equal to the thermal penetration depth. For now, we'll take the radius of a patch to be equal to the penetration depth, which we'll also consider to be identical to the coherence length for the fluctuations. I am being very sloppy with factors of two here. It is the difference between doing something in a few minutes and doing something in a few hours. To get the details right, I would have to do some very careful integrations. Instead, we'll be almost right.

We know how to write the thermal penetration depth so we can write the phase fluctuations as shown. Here, all of the coherence lengths or radii have been replaced with the thermal penetration depth.

In the fiber, the optical field is concentrated in the vicinity of the fiber axis. This region is called the mode field and is characterized by the mode-field radius, a_{mf}. Of course, the physical fiber is considerably larger but the majority of the energy is concentrated in this central region by index grading.

At a particular frequency, we could sketch the arrangement of coherent patches – the size of the patches depends on frequency through the thermal penetration depth. In the diagram shown, the mode field cross-section contains ten or so patches. The ultimate effect on the optical signal is an average over the fluctuations produced by these ten patches. We don't need to consider details smaller than the patches because the thermal fluctuations are coherent on smaller scales. So much for the effects across the mode field. As the light travels along the fiber, the effects of the patches accumulate. In particular, the mean-square fluctuation increases linearly with length along the fiber.

Even if you haven't followed the math, you should notice that something different is probably going to happen if the thermal penetration depth equals the mode-field radius. I'm sure you all were thinking that, right? Good.

[Transparency III-13]

The sketch on the previous slide really represents the high-frequency limit where the thermal penetration depth is small compared to the mode-field radius. In that case, we averaged the fluctuations over the patches in the mode field. The averaging reduces the mean-square fluctuation as the ratio of the area of a patch to the area of the mode field. As the frequency increases still further, the fluctuation patches grow ever smaller and the mean-square fluctuation is further reduced. Note also that the mean-square fluctuation is proportional to the fiber length.

On the other hand, if the thermal penetration depth is much larger than the mode-field radius, then there is no averaging over the mode-field cross-section. The entire cross-section shows the same fluctuation. In this regime, the fluctuations are independent of frequency (but still proportional to fiber length).

[Transparency III-14]

Since there's been so much interest in optical fiber systems, it's relatively easy to find the parameters required to calculate the actual fluctuations in phase. As an example, let's consider a fiber length of 100 m. For this single-mode fiber, the thermal penetration depth is equal to the mode-field radius at about 50 kHz. Below this frequency, the mean-square phase fluctuation (the spectral density, that is) is independent of frequency and is a bit more than one microradian per root hertz. Above this frequency, the spectral density of the phase fluctuations drops.

This is very important for sensors based on optical fibers. It is frequently claimed that it is only necessary to wind more fiber on a particular sensor to improve its performance (lower its noise floor, for example). Even if the noise floor were independent of fiber length, at some point the responsivity would be high enough that intrinsic noise in the sensor mechanical structure would take over as the dominant noise source and further increase in fiber length would only increase the cost, not the performance. However, especially for a sensor designed for signal frequencies below 50 kHz, the phase fluctuations introduced by thermal agitation in the fiber itself increase linearly with fiber length. For 100 meters of fiber, we are up to 1½ microradians per root hertz. This is at or near the limiting noise floor for a very good demodulator.

If we use 1000 meters of fiber, we are not going to have to buy a very expensive detector, because the fiber will limit the performance well before a good demodulator would. Issues of noise are extremely important in fiber systems but the lessons should not be restricted to fiber sensors. *Any time a "very sensitive" detection mechanism is "discovered" the questions of limits to performance and self-noise should be among the first questions asked.*

[Transparency III-15]

I'd like to end the discussion of noise in fiber-optic systems by going through the analysis for shot noise. This is a popular calculation but it's actually not terribly important in most cases. Usually, the actual noise floor of an optical-fiber sensor is well above the floor predicted by the shot-noise calculation. It's worth discussing, though, because it does give a performance limit and, in a few instances, the shot-noise limit has been approached.

The basic photodetector response can be written as the photodetector current being proportional to the incident optical power. The response factor, B, has the units of amps per watt (or the more "tasteful" microamps per microwatt). The fundamental interferometer response produces a sine-wave like output as a function of the phase difference between the two beams. The fringe visibility is a measure of how complete the interference is. If the two beams are perfectly balanced, then, for destructive interference, the resulting intensity will be zero. This perfect balance is equivalent to a fringe visibility of one. The maximum sensitivity is obtained at the point where the response function is changing most rapidly. If we calculate the rate of change of photodetector current with phase diffference at this maximum point, we find that it is equal to the fringe visibility times B times the incident optical power.

[Transparency III-16]

The mean-square value of the current fluctuations is equal to the standard shot-noise form based on the average photodetector current – 2 times q times I_0. We could have also written an equivalent expression for the mean-square fluctuations in optical power as 2 times hf_0 times the average incident optical power, $\mathbf{P}_0$. The quantity, hf_0, is Plank's constant times the optical frequency and it is the energy per photon. Dividing the second form by B^2 gives another expression for the mean-square current fluctuations.

From the relationship between the phase and the photodetector current, we can convert the current-noise expression into a phase-noise expression. The spectral density of the phase fluctuations is 2 times Plank's constant times the speed of light divided by the quantity, fringe visibility squared times the optical wavelength times the incident optical power. For any particular system, we can now calculate the shot-noise limit. Notice that the phase noise is inversely proportional to the incident optical power so, if shot noise is significant, it is likely to be so in systems operating with very low optical power. For an interferometer operating at 630 nanometers and only 10 microwatts of optical power, the shot-noise floor is well below the limit achievable by practical demodulators.

People have, in fact, reached shot-noise limits in fiber systems, but it takes hard work, and few people succeed. The other noise mechanisms I've discussed are generally more important in fiber systems, but everyone working with such systems should appreciate the roots of shot noise. Please don't ever believe that once you've calculated the shot-noise limit you've found the actual noise floor, though.

[Transparency III-17]

Not all noise problems can be described in terms of thermal or shot noise. Some noise mechanisms are not as straightforward. However, most if not all noise mechanisms involve interesting physics. Now, I would like to talk about a fascinating problem regarding noise in ferroelectric materials. The noise process that I'll describe is not one that can be treated as a fundamentally thermal or shot process.

Ferroelectric materials are those materials that we usually call piezoelectric materials. The class of piezoelectric materials is a broader class. All piezoelectric materials show evidence of linear charge "generation" under stress or linear dimensional change under application of an electric field. Ferroelectric materials are piezoelectric but they are a particular variety of piezoelectric material. These materials have a permanent electric dipole moment that can be manipulated. Under the application of a sufficiently high field (perhaps also at elevated temperature), the direction of permanent moment can be reversed. This is the operation that is normally called "poling." One of the consequences of this "flippable" dipole moment is that these ferroelectric materials are more strongly piezoelectric than ordinary piezoelectric materials like quartz. In addition, they are also pyroelectric – that is, they also respond to application of heat.

The buzzer disk in your watch is made from a ferroelectric ceramic. Ultrasonic humidifiers are driven by ferroelectric transducers. Many types of hydrophones, microphones, and accelerometers are made from ferroelectric materials.

Besides having large piezoelectric coefficients, these materials have other very interesting properties. In 1957 some folk at Penn State published an intriguing set of measurements on the ferroelectric material, barium titanate. They took a small piece of it and applied stress to it. The stress was applied suddenly – a step-function in stress. After a half-hour or so, they released the stress suddenly. During this process, they measured the strain produced.

In an ordinary elastic material, the strain is directly proportional to the applied stress without any delay in response. However, this material does not behave that way. The graph in this slide shows the behavior of a sample at several temperatures. The horizontal (time) scale starts at the onset of the stress. The applied stress is 24 MPa. At room temperature, the initial strain is about 270 microstrain. Over a period of about 20 minutes, the strain creeps slowly up to about 360 microstrain. At a somewhat higher temperature, the creep is still evident but it reaches equilibrium more quickly.

At still higher temperature, the strain is quite a bit lower but now there is no evidence of creep. This last temperature is just above the Curie point for this material, which I will talk a bit more about. Below this temperature, the process is very interesting because of the time required to reach the equilibrium strain. If you see a process like this, you might suspect that there is a noise process with a similar time scale lurking in the background and, in fact, there does seem to be a noise process associated with this mechanical creep.

MR. GLADDEN: On that slide, is the stress released at about 30 minutes?

DR. GABRIELSON: Yes, at 30 minutes they released the stress. After the stress is released you see a relaxation back the other way. This is a reversible creep. It did not happen just once. Each time they applied the stress, the strain followed the same curves.

[Transparency III-18]

This is a molecular structure that is representative of ferroelectric materials. In particular, this is the unit cell for barium titanate. Above a certain temperature, the structure is cubic with no preferred orientation. The blue atoms (the largest ones) are oxygen atoms; the red atoms (at the corners of the cube) are barium; and the little black atom in the middle is the titanium atom.

If I take a slice through the center of the unit cell, I see a titanium atom surrounded symmetrically by oxygen atoms. The titanium atom has only one stable configuration and there is no net polarization (charge asymmetry). As the temperature decreases, there is a rather sudden transition at the Curie temperature. Here, the crystal structure changes from cubic to tetragonal. (Tetragonal just means that the cube has been stretched in one dimension; all the angles are still right angles.)

Now the central slice shows that two of the oxygen atoms are much closer to the titanium atom than the other two. As a consequence, there are two stable locations for the titanium atom. In either stable location, there is net charge asymmetry and, therefore, a permanent dipole moment. If we apply a strong enough electric field, we can flip the titanium atom into the other stable state. This is the essence of poling a ferroelectric. Furthermore, the "permanent" moment produced by the "flippable" charge asymmetry give the ferroelectrics most of their interesting properties.

Even when poled by a strong electric field, the moments are not completely aligned throughout the material. Domains form. In each domain, the unit cells are poled in the same direction. In an adjacent domain, the direction of poling might be at 90° or 180° with respect to its neighbor. The boundaries between domains are called walls and they may be either 90° walls or 180° walls depending on the difference in orientation across the wall. Of course, when the material is poled there is a strong tendency for the domains to be oriented in the poling direction but the thermal energy is sufficient that there are still many domains perpendicular and opposite to the poling direction. Under stress, the walls "move." Actually, unit cells near the wall boundary re-orient. Many people believe it is the migration of the 90° walls that control the physical properties of the material.

As I said, if you apply a stress, the walls move, but it takes time for the walls to move (or, equivalently, for the domains to reorient). What happens when the dipole moment of a unit cell changes direction? It produces a change in the charge distribution and that is sensed as noise.

[Transparency III-19]

This is quite different from microcracking because this is a reversible process. Some years ago, people noticed that some kinds of hydrophones produced noise for 30 minutes to an hour after they were dropped into the water. Because the heat capacity of the water is large and there was a strong flow past the hydrophone as it descended to depth, the temperature of the hydrophone reaches equilibrium much faster than 30 minutes. (Originally, the noise was thought to be a pyroelectric effect but the time scale just doesn't make sense.)

Here are some measurements I made shortly after this problem was discovered. I made some structures of piezoceramic material bonded to aluminum. Some structures had ceramic on only one side; others had ceramic on both sides of the aluminum substrate.

If the temperature changes, the difference in coefficient expansion between aluminum and the ceramic results in a large stress on the ceramic. This is a better way of stressing the ceramic than pushing on the structure; if you try to stress the ceramic by pushing on the structure, you introduce so much extraneous vibration that the stress-induced ceramic noise is masked.

Not much of a temperature change is required. For these measurements, I used a temperature change of 10° C. After the temperature change, there is a significant amount of noise bursts to 10 minutes or so for the single-sided structure. Initially, the bursts are strong and very close together. As time goes on, the level and frequency of the bursts decrease. Interestingly enough, in the double-sided structure the bursting noise takes considerably longer to decay. You can see noticeable bursting events almost two hours after application of the temperature change. The induced stress is substantially different between the two types of structures so it is not surprising that the noise characteristics would be different. Unfortunately, from the standpoint of signal response, the double-sided structure is better.

The noise generation is repeatable. If I reverse the temperature change, I see the same sort of noise production. It is not a cracking phenomenon because it is repeatable. Although it remains to be proven, it appears that this noise phenomenon is closely linked with the anelastic behavior of the material through the mechanism of domain wall motion. In any event, this is another example of a physical process into which we gain some insight through study of its noise.

[Transparency III-20]

For the last subject I would examine the impact of feedback systems on noise. This subject is discussed in the literature at least as far back as the 1930's. A little more recently, Dr. Keolian published a paper about a nice demonstration apparatus for feedback-controlled systems. This is a very instructive demonstration to reproduce. Dr. Keolian describes the construction and operation and he also discusses some of the practical issues associated with the system.

I am going to skip the practical issues. Let's consider a simple mass-spring oscillator with damping. (I've left out the noise-force generator but you know how to put that in.) If you put two magnets on the mass, you can drive the mass with one magnet (and a coil) and detect its motion with the other. If I amplify the detector output as shown, then the output voltage is some constant times the velocity of the mass.

If I take that voltage and pass it through a standard voltage-to-current converter, I can produce an actuation force that is proportional to the mass velocity. Now, my schematic drawing is sloppy. There would also be direct electromagnetic coupling from the actuation coil to the detection coil. Dr. Keolian came up with a nice solution for that problem but you'll have to read the article to see how.

[Transparency III-21]

Let's consider the system without feedback first. Some applied force produces an output voltage. The block **A** represents the conversion between velocity and voltage. The differential equation is straightforward. The force equals the mass times the acceleration plus the resistance times the velocity plus the spring constant times the displacement. Of course, I should consider noise as well.

If I connect the output back to the input through block **B**, which represents the conversion from voltage to force, then I've added another term to the force. By combining terms, I discover that the damping now depends on the original mechanical damping minus this A-times-B term. The damping can be increased or decreased. In fact, if the AB term is large enough (and positive), the damping becomes negative and the system will oscillate. This is standard control theory but what happens to the noise? We've argued that the noise is linked strongly to the damping. What happens if we change the damping not by changing the physical loss mechanism but by adding feedback in such a way as to modify the apparent damping?

[Transparency III-22]

As always, what's really important is the signal-to-noise ratio. There is a very nice theorem that we will use to simplify our argument. For a linear system, the signal-to-noise ratio at the output terminal (this is, at the load impedance) does not depend on the value of the load impedance.

Often, feedback loops are "broken" when the load impedance goes to zero and, if the feedback loop is gone, the analysis is usually considerably easier. We can either use the usual voltage ratios for SNR and let the impedance approach zero or let the impedance be zero and calculate the SNR in terms of the signal and noise currents.

[Transparency III-23]

Now, I can certainly change the effective *Q* of the system with feedback. I can make the *Q* either higher or lower, depending on the sign of the *AB* term. If I decrease the effective resistance, can I reduce the noise of the system?

Let's use the theorem. If I calculate the signal-to-noise ratio for the system, *A*, without feedback, then add feedback, *B*, what happens? If I set the load impedance to zero, the feedback path is broken. The signal-to-noise is that of the original system *PLUS* whatever is contributed by resistances or electronic noise in the *B* path. It turns out that I can never improve the signal-to-noise ratio by adding feedback if the dominant noise is in system *A* or in the input to system *A*. (If the noise is dominated by something *after* the AB feedback system, then I may be able to use feedback to increase the overall signal level and improve the SNR until I've also raised the internal noise high enough to become the new dominant source.)

If I'm very careful, sometimes I can make the noise added by the B path small and accomplish something. In some cases, this is useful.

[Transparency III-24]

In or around the 1950's, Mylatz performed a nice set of experiments to show how both signal and noise change with feedback. There are two strategies to consider. Design the *Q* of the system to be very small (to get a flat response, for example) but then add feedback to increase the sensitivity of the system. The intrinsic noise is determined by the real, physical damping mechanism and, if the *Q* is small, the damping associated noise will be large. The signal-to-noise ratio is not going to be very good to begin with; adding feedback is only going to make it worse. The reason the feedback-induced damping does not control the intrinsic noise is that the feedback-induced damping is not equilibrium loss. The feedback is precisely (within the limits of the electronics, that is) coherent with the signal *AND* the noise so it is not a process that contributes additional fluctuation proportional to the loop gain (i.e., to the artificial damping). The natural damping reflects a process that is in thermal equilibrium with the environment, so that controls the thermal noise.

Another strategy is to design the mechanical Q of the system to be naturally very high. That produces a good signal-to-noise ratio with respect to intrinsic thermal noise. Then we can add feedback to flatten the response.

In the process, of course, noise is added by the feedback loop. However, if I've designed the feedback carefully I still have better SNR than if I had made the response flat just by making the natural damping large. It is not easy. I have tried it in a number of cases and it is difficult to get much of an advantage, but you can sometimes improve the SNR significantly.

In some circles, this is called "cold damping." There was even a paper in the *European Journal of Physics* this year (2000) that discusses cold damping. As with many aspects of noise, if you understand the fundamentals, then you can make some wise decisions about design strategies.

We've been embroiled in details in this last hour. Let's summarize. We've discussed the very fundamental thermal-noise mechanism – the noise associated with temperature fluctuations. This variety of noise shows up in many interesting places and it is fairly easy to do calculations. It is important in fiber-optic sensors with relatively long fiber lengths. It is important in microfabricated sensors that have very small moving parts. You ignore such a fundamental mechanism at your own peril. You may postulate wonderful performance and fail to achieve it if you ignore noise.

We also saw that there is a basic relationship between some quantities that we associate with purely deterministic processes and quantities that are associated with noise. I measured the real part of an impedance by making a noise measurement and I calculated the imaginary part of the impedance from the real part using the Kramers-Kronig relationships.

We talked in some detail about molecular collisions and conduction electrons in metals and semiconductors. We saw a nice analogy between these two cases and, hopefully, learned something about shot noise in the process.

Then we examined some special applications in which the principles discussed in the first two hours were extended in several ways. Thank you very much for your kind attention.

A Selective and Terribly Inadequate Bibliography of Noise

Basics of Fluctuations

R. Kubo, "The fluctuation-dissipation theorem," Reports on Progress in Physics, Vol. 29, Part I, A. C. Strickland, ed., 1966.

C. Kittel, *Elementary Statistical Physics*, Wiley, New York, 1958.

L. Landau and E. Lifshitz, *Statistical Physics, Part 1*, 3rd Ed., Butterworth-Heinemann, Oxford, 1980.

A. Pippard, *Response and Stability*, Cambridge University Press, Cambridge, 1985.

F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York, 1965.

Thermal Noise

G. Uhlenbeck and S. Goudsmit, "A problem in Brownian motion," Phys. Rev. **34**, 145-151, 1929.

H. Ellis and E. Moullin, "Measurement of Boltzmann's constant by means of the fluctuations of electron pressure in a conductor," Proc. Cambridge Phil. Soc. **28**, 386-402, July 30, 1932.

J. Johnson, "Thermal agitation of electricity in conductors," Phys. Rev. **32**, 97-109, 1928.

H. Nyquist, "Thermal agitation of electric charge in conductors," Phys. Rev. **32**, 110-113, 1928.

W. Moll and H. Burger, "The sensitivity of a galvanometer and its amplification," Phil. Mag. **50**, Series 6, No. 297, 626-631, 1925.

C. Byrne, "Instrument noise in seismometers," Bulletin of the Seismological Society of America **51**, 69-84, January, 1961.

Shot Noise

A. Hull and N. Williams, "Determination of elementary charge E from measurements of shot effect," Phys. Rev. **25**, 147-173, 1925.

W. Schottky, "Small-shot effect and flicker effect," Phys. Rev., **28**, 74-103, 1926.

R. Oliver, "Thermal and quantum noise," Proc. IEEE **53**, 436-454, 1965.

Transistor Noise

Van der Ziel, "Theory of shot noise in junction diodes and junction transistors," Proc. IRE **43**, 1639-1646, 1955.

W. Guggenbuehl and M. Strutt, "Theory and experiments on shot noise in semiconductor junction diodes and transistors," Proc. IRE **45**, 839-854, 1957.

K. Lonngren, *Introduction to Physical Electronics*, Allyn and Bacon, Boston, 1988.

P. Gray, *Introduction to Electronics*, Wiley, 1967.

Fiber-Optic Sensor Noise

K. Wanser, "Fundamental phase noise limit in optical fibres due to temperature fluctuations," Electronics Letters **28**, 53-54, January 2, 1992.

W. Glenn, "Noise in interferometric optical systems: an optical Nyquist theorem," IEEE J. Quantum Electronics **QE-25**, 1218-1224, 1989.

Feedback and Noise

J. Milatz, J. van Zolingen, and B. van Iperen, "The reduction in the Brownian motion of electrometers," Physica **19**, 195-207, 1953.

M. Usher, I. Buckner, and R. Burch, "A miniature wideband horizontal-component feedback seismometer," J. Physics E: Sci. Instruments **10**, 1253-1260, 1977.

R. Keolian, "A simple harmonic oscillator teaching apparatus with active velocity feedback," J. Acoust. Soc. Am. **99**, 1260-1261, 1996.

Kramers-Kronig Relations

J. Jackson, *Classical Electrodynamics*, 2nd Ed., Wiley, New York, 1975.

Ferroelectric Materials

C. Kittel, *Introduction to Solid State Physics*, 7th Ed., Wiley, New York, 1996.

RESONANT ULTRASOUND SPECTROSCOPY

Albert Migliori
Los Alamos National Laboratory

DR. MIGLIORI: Before I start the talk, this is a picture of me being shot through a glove box window.

[Transparency]

I am behind two sets of chain-link fences with razor wire on the top and there is about a 10-foot separation between the two fences. In-between the fences are microwave and infrared personnel detectors. Looking up over the fences all around are guys with 50 caliber and M60 machine guns, pointing them out through little ports about this big.

After you have gotten 2 million security clearances, random drug testing, et cetera, you can go through the first set of fences. Then another 2 million or 3 million pieces of paperwork and you can go into the changing room, strip, put on what Los Alamos calls anti-C, for anti-contamination, clothing.

Then you head in through another set of guards with M16s and all sorts of stuff and they lock you up between two little gates, take your badge, look at you, read your palm print and then, if they like it, they let you inside a vault door, which is about an 18-inch-thick piece of solid steel inside a building within a building at the plutonium facility at Los Alamos.

This building actually has a set of airtight seals that can close the entire building, and they know that there are no leaks, because they check this. The roof is a whole bunch of stressed concrete and stuff, and they claim you can crash a 747 nose-end into it and it will not collapse the building.

You go inside in there and you have one of your resonant ultrasound laboratories, and in that laboratory we have access to plutonium. In fact, some of the measurements that motivate my recent work have been the first measurements of elastic properties of plutonium since 1973.

Because of the extensive security and other regulations required to get in here, I cannot use postdocs, you just cannot get them clearances in time, so I actually have to do work myself. (laughter)

Right over here in my hand -- at this point here -- is a 14.something-gram chunk of plutonium 239, which is weapons grade plutonium. That chunk of plutonium was removed from a thermonuclear weapon from what is called "the pit," or the fission trigger, from a thermonuclear weapon.

It was given to us to purify and then make resonant ultrasound samples of.

I think this experiment was probably responsible for the destruction of one thermonuclear weapon. [Computer slide]

Part of the reason for mentioning all of this is that I just have to make a comment about the recent events at Los Alamos. I do not know of anybody I work with who would be so unbelievably idiotic as to download secrets from a classified to an unclassified computer, or walk out the door with a bunch of hard drives. I do not know what is going on here, but I think it is a very rare event. They have been exceedingly careful in all the aspects of working with these very dangerous materials, both from the health standpoint and from the political and weapons standpoints.

You can see behind me here we have a continuous air monitor, there are fire alarms in the glove box, pressure monitors in the glove box. The room, security, everything, I feel very comfortable working with this stuff at Los Alamos.

Another minor point here is that this sample, by the way, I can feel the heat from its self-radioactive decay in my glove. [Computer slide]

I am going to shut this stuff off and continue.

The fire at Los Alamos, by the way, completely surrounded the building in which these experiments were done and there were flames on all four sides of it. In the process, the laboratory was shut down for two weeks.

That two weeks turned out to be precisely the two weeks before Libby required me to deliver viewgraphs for this presentation.

I had wanted to try something distinctly different this year, so although I did not succeed completely in getting the viewgraphs in the form I wanted -- you will see that in the middle of the talk, where I actually had to copy some preprints that I had prepared for this for another reason -- but I wanted to do this year is provide you with a set of preprints in your notes, rather than viewgraphs, which, when you go back and read them, for those of you who find that this may be something you care to work with in the future, you will find a complete set of self-contained publications there, with references.

I think, in addition, because of the detail that is in those, I will not cover every one of those at all; guaranteed that I will not go through all the viewgraphs you have in your lecture notes. I also guarantee that if you have trouble sleeping, they are an instant cure for insomnia later.

[Transparency]

I will begin by giving away the entire talk in the first part of this. Then I will cover bits and pieces of details on -- well, let's see, classical mechanics, solid-state physics, acoustics, statistical mechanics, electronics design, maybe a little relativity; we will try to get through all the basics.

I will give you some details, again, which will be a little dry at the beginning, on how the method works. But then something very important that I want you to take away with you is the importance of ultrasonic measurements in solid-state physics.

There are very, very powerful motivations for making sound speed measurements in solids. Often glibly, those are thrown off as the elastic stiffness that controls the sound speed is a second derivative of free energy with respect to strain.

What I am going to do in the middle part of the talk is to, point by point, precisely and in detail, show you exactly why making sound-speed measurements can be some of the very most important measurements in all of solid-state physics. It is a shame that they are hard to do, so they are not used as frequently as they should be, and I will show you how that gets famous theorists in trouble. Then I will show you some results.

This year I am going to present bits and pieces of electronic design. The talk this morning was perfect, because I have kind of spent the last several years implementing instrumentation systems in which I aspire to do things the way Tom described this morning, but I will show you where you cannot and how we made some compromises that still result in stunning homemade black boxes, and I think you will like that.

[Transparency]

Orson Anderson at UCLA is really the father of resonant ultrasound spectroscopy. He began this in the late 1960s with his postdocs and students at Bell Labs and then used the technique to measure the speeds of sounds in lunar finds. These were small spherical samples of rocks, spherical because of impact and heating on meteor strikes, that were returned by the Apollo missions from the moon.

Those samples had been very carefully vacuum-shipped from the moon. They took great pains not to contaminate them. Orson really developed resonant ultrasound spectroscopy -- at least everything that he did then we do now (it may be more difficult, more complicated) -- but all the bits and pieces were there.

He used this technique of measuring mechanical resonances of small objects to extract the elastic moduli to study the stiffnesses of rocks that appeared on the moon. These rocks also appear on the earth and what he found was that the speeds of sound in these rocks were anomalously low and, in fact, if you look at the sound speeds in kilometers per second here for a couple of lunar rock samples, you will see that they fall very nicely amongst the speeds of sound of various cheeses. (laughter)

Not to be deterred, and with very little analysis, because Orson had already figured out the answer, he, of course, published this in Science.

What had happened was that these lunar finds had been vacuum-cracked on the moon, so instead of being a homogeneous uncracked solid sample of rock they had microcracks. They still looked like a homogeneous medium to a long wavelength sound wave but they were much softer.

As soon as he brought them out and began to let moisture heal these cracks, the sound speeds came right back up to the normal Earth speed, so they discovered something. He did put out this Christmas card.

"SEASONS GREETINGS
It brightens the spirit
In times like these,
To know the Moon
Is made of cheese."

By the way, Orson Anderson, in World War II, was a fighter pilot who also had a cartoon strip that he regularly published, and he liked to do these sorts of things.

I really have to thank Orson, but I did not get to thank him when I began this work. In fact, when high-temperature superconductors came out, I realized that we were not going to get large enough samples to do conventional pulse-echo ultrasound measurements on them, so I began to use and set up experiments to use resonances to extract elastic mmodulus information.

We did, indeed, succeed in putting together all of the pieces and had published one or two things, when one day Orson Anderson walked into my lab in Los Alamos, he had come out from UCLA, and he said that he wanted to have me help him do some electronics. I said, "Orson, what for? Glad to meet you, what do you want to do with it?"

He said, "I'm measuring the resonances of small samples." I said, "Oh, so am I." He says, "We're using it to get elastic constants." "Oh, I'm doing that, too." As it turns out, I guess Jay Maynard quoted me in Physics Today, "Six months in the lab can save you a day in the library," and I fell victim to that.

It was a good thing, too, because we did take somewhat different approaches to the computation and the electronics development and, in the end, those approaches made this technique quite usable for almost anybody who really needs to make sound-speed measurements.

[Transparency]

The basic principle is that the measurement of the resonances of small samples with accurate geometry can get you elastic moduli. This is a small sample of steel inside some half-inch diameter assemblies that are used as transducers, and we will talk about every last detail that I go through at the beginning here.

[Transparency]

Having set up a transducer assembly, we have to acquire data and, as you noticed in the other picture, the sample is making point contact. This is because we want to drive all the resonances and, therefore, we want to touch the sample at a low symmetry point so that we excite all the modes.

That means that we have dry point contact, we do not have to glue transducers on to the sample and with that dry point contact comes very weak electrical coupling, so great attention has to be paid to the preamplifiers, the detection, and the data-acquisition systems.

[Transparency]

Unfortunately, solids behave completely differently from liquids and gases when sound propagates through them, so you have to use a much more sophisticated analysis technique to extract the mechanical properties from the resonances; that is, the mechanical resonances of a solid are not simple functions of 1D parameters. The reason is, when I squeeze a solid, it bulges, and it bulges so that the volume is not preserved.

The analysis now becomes quite sophisticated, but we are able to calculate normal modes for such objects.

[Transparency]

This is an example of some of them for a solid that is orthorhombic or higher in symmetry, and we will go over that in a little bit, too.

[Transparency]

Each of these modes is either symmetric or anti-symmetric across three separate planes. We can take the data, we understand the mechanics problem, and we can compute the normal modes, and we can even go backward, so that when we measure resonances -- I will keep popping up plutonium here, just because I have been working on it recently, but also because it turns out to be a very average sample for resonant ultrasound spectroscopy. It is not the very best material to do it in, but it is not the very worst, and we get very nice data.

The systems we are using now are optimized for all sorts of signal dynamic range and noise parameters, but you can see that on a small -- this happened to be a 1.43-gm sample -- you can see that we have signal-to-noise and full-phase sensitivity, so we can acquire those resonances quite nicely.

[Transparency]

The resonances produce elastic moduli (and I will go over that a little bit, too), but there is an approximate bottom line to this and that is that this technique probably -- not probably, it does have the highest absolute accuracy of any routine elastic modulus measurement technique.

This has been tested against single-crystal silicon in a round robin between us and other laboratories, including NIST -- Hassel Ledbetter and the late Chris Fortunko. Here is an example of a relatively good sample -- still not the best and, in fact, I am not going to show you the best, because you just would not believe it.

No, that is not true. The reason I am not going to show you the best is because my office was locked for two weeks and I barely had time to get the viewgraphs out. They would not let us in.

This is an example of why people use resonant ultrasound and not other techniques when they really need elastic moduli. It turns out that the bulk modulus stiffness on hydrostatic compression is independent of whether I have a single crystal of a material or a bunch of polycrystalline grains put together in an ordinary commercially processed piece of material.

In this case we were looking at beryllium, which has the second highest sound speed, diamond being the first. It also has an extremely curious property that when you squeeze it, it does not bulge.

We were measuring single crystals of beryllium here and a polycrystalline material, both of these high purity. Although this looks like a viewgraph you should never show, I am going to walk you through it a little bit and then we will come back to it later.

What I want you to notice is that we can measure resonances (these are in megahertz), we can compute elastic constants by fitting the five independent moduli for the single crystal or the two independent moduli for the polycrystal and we come out with error bars in the ballpark of these sorts of numbers consistently, but those error bars are computed on the basis of statistical analysis of the data.

There should be a check, and the check is that these two should have the same bulk modulus, and here you go, so it does quite well. It is also quite sensitive for various bits and pieces of solid-state physics.

[Transparency]

Here is a plot of the deviations in a plutonium, not the beryllium sample, but this is very typical. These are the percent errors that a fitted set of elastic moduli produce when you compute the frequencies that you should have measured and compare them to those that you did measure. The technique works great. It lets you see lots and lots of marvelous stuff.

[Transparency]

This, for example, is a superconducting transition in niobium at around 9 Kelvin. There is a break in slope of the shear modulus. This distance between here and here is about 3 ppm, so we are watching sound-speed changes at the 3-ppm level as the electronic distribution shifts from ordinary metal to superconductor in niobium.

At low temperatures, as we all know, everything gets nice, so this particular sample had a mechanical Q of 109,000, which we can measure to about 1% -- right, pretty nice, you do not see that in gases.

[Transparency]

The technique also has commercial applications, which I have kept away from in the last few years, relating to quality control. This was the first launch of the Trident II -- those are Logan's boys. (laughter)

Here it is, going about Mach 1, straight into the ocean.

There are applications to quality control as well as to the study of science. One way to see how that can happen is to remember that we are easily able to see parts-per-million stuff. Here is a ball bearing that is a little bit out of round, so that one of its normal modes -- like this -- if it were perfectly round, this normal mode and this normal mode would have the same frequency, but if it is a little off in roundness and the material is isotropic, they split, you break the degeneracy.

Here is a split in those two modes. It turns out that for spheres there are odd numbers. When they split, it is sort of 2L+1, so you have to have an odd number of modes, and it turns out there are actually two modes under this peak and one over here.

The 700-ppm shift between the modes can be quantitatively connected to an error in manufacturing, how out of round this ball is. In the end, this technique has become important at some leve

[Transparency]

This is a book by some obscure authors on this subject. This is a commercial system that I actually did the electronics design for that does, in a nice and sophisticated way, the acquisition of resonance data for small solid samples.

DR. HARGROVE: What is the scale on that?

DR. MIGLIORI: Let's see, this book must be 8 x 10, and they are all on the same scale, so actual size.

That is what I am going to tell you about. To do that, I am going to start with some viewgraphs that have been taken from that book. The intent here is to just warm up by examining the properties of resonances, so I will not follow it page by page, as you have it, but I will pick and choose things. So here we go.

[Transparency]

We are looking at resonances; I need masses and springs. Having played with the measurement of solids for longer than I can remember, I know that for all of the experiments I do there is no such thing as a rigid wall, so I drew this resonator without the proverbial brick wall that you attach the springs to, so this is floating in free space and it has two masses -- like this -- and a damping piston and a spring.

There are a lot of interesting things to note about this when you start to look at center frequencies of resonances with Q's of 10^4 or 10^5, where the resonances are very sharp, and I will chat about that for a minute here.

[Transparency]

You are all experts in this. Right now we are going to set the equations of motion up so that we have inertial pieces, the ma part, F=ma, damping springs. There is F=ma with the factors of two in my own notation here. We are going to assume harmonic motion and you know we are going to get complex frequencies from that eigenvalue equation.

[Transparency]

This has been covered so many times, but we are now in a realm where we can actually see some of these shifts; that is, the Q's can be typically so high that I can see small effects.

For example, we can see that the real part of the solution to that mass-spring equation is slightly shifted by the dissipation -- I think everybody has discussed that -- and there is an imaginary piece that kicks in. You notice that the change in the real part of the frequency is quadratic in the dissipation while the imaginary part is linear in the dissipation. We can define sort of a canonical resonant frequency for this thing of two times the spring constant over the mass, just exactly the sort of thing you have already seen.

[Transparency]

But there are some twists to it that have to do with plotting out the response function. We cannot see, mostly, the effects of thermal noise in these systems, so that even though Tom, two years ago, did his very nice viewgraph with the thermal-noise generator, the problem with these systems is that we detect building vibration and everything else at levels much higher than the thermal-noise background for the mechanical system, so I am going to ignore it, roughly.

But I am not going to ignore the fact that we are measuring the complex amplitude and, in fact, the electronics I now use preserve the full complex response so that I know just what is going on. It is, of course, approximately Lorentzian (it is not exactly, because it is not a single isolated resonance).

[Transparency]

Remember, in these mechanical systems there will be many resonances, so the tail of one overlaps another, and all of those are relatively small effects, but there is a particularly interesting and not-so-small effect under some circumstances. This relates to the termination of the center frequency of the resonance.

How do you know what the center frequency of these resonators is? There is the normal counting problem in measuring frequencies; that is, at the very most fundamental level there is a one-count error in your measurement and if you try to beat it, you will not, because then noise will kick in.

For example, if I take a pulse train and I start my discriminator at the beginning of the train and I wait for 10^6 cycles and I count how many I have in there and then I stop it, I have to stop it at some point. If I stop it right on the transition, I might get a count error of one in that total, so I waited one second, I counted 10^6 things, I get a megahertz, but sometimes I get 10^6 plus one things going on, and sometimes I might get 10^6 minus one going on, and you cannot beat that one, because noise will force you to be unable to determine precisely where that count is, even if you have a very clean system, so there is an intrinsic problem in measuring frequency of about one count. That is not so bad, because I can wait as long as I want.

There is another set of problems in a system that appears to be essentially noise-free and, for the purposes of this discussion, we will assume it to be noise-free, and that is that the frequency is not quite well-defined in a damped problem with a complex eigenvalue.

For example, the displacement is a maximum when the frequency equals R $\underline{\text{omega}}_0$ the square to 2k/M times 1 -1/2Q_2 to the ½. That is the biggest displacement the oscillator sees. That is not quite the real part of the eigenvalue when we solve that equation -- notice that this has a 4 here, and that has a 2 there -- and it is not the frequency at which the displacement is exactly out of phase with the force, which is $\underline{\text{omega}}_0$.

There are essentially three choices here for what the center frequency of that resonator is and they all differ by order $1/Q^2$. If Q is 100 and I can see parts-per-million things, I can actually see that sometimes.

There is another physics aspect to it, and that is that, intrinsically, the elastic response of a solid is defined for a dissipation-free solid. As I begin to increase dissipation, that is, reduce the Q, the meaning of elastic constants starts to change; that is, mechanical systems start to open up the response into a hysteresis loop and it is not quite so clear any more what an elastic constant means because of that hysteresis and dissipation.

That is reflected in this problem here. I just wanted to show you that briefly to make clear that if we solve a linear problem with no dissipation and then go and measure, we have to take some pains to make sure that either we can ignore those problems or we know what they are.

[Transparency]

Here is a response of a typical resonance from those. Here is the amplitude. Here is the real part and the imaginary part. Another thing to notice, which we actually make use of, is that if I simply plot the real part of the response of this resonator, it is apparently narrower, much narrower, than either the imaginary part or the actual displacement.

That comes in handy when we are trying to separate closely spaced resonances.

MR. APOSTOLOU: Which one of the three would you choose to have the resolution to the speakers to actually get the elastic constants? Which relates closer to --

DR. MIGLIORI: The one that relates most closely is the one where you are in phase with the force. It is the zero crossing of the imaginary part, if you will, because that is the one that is the square to 2k/M. If the noise properties of the system are good enough, that is fine. Of course, if the Q is 10^5, then Q_2 is 10^{10} and we are home free, we do not care.

[Transparency]

We are dealing with solids here and stress waves in solids have lots of stuff going on, including shear waves, that the standard liquid or gaseous acoustics problem does not have to address. That becomes very painful, extremely painful.

There is this interesting question as to when a material is a liquid and when it is a solid, which I think has a lot to do with the viscous penetration depth. For example, is the viscous penetration depth is a few light ears, it is probably a solid, but if it is a few millimeters it is probably a fluid.

Just think about window glass, for example, with time constants for drooping of 10^3 years. The stuff behaves, for all intents and purposes, like a liquid over very long time scales but, like a solid, over short ones. There are some complications in there, but we are not going to worry about those, because anything that shows any signs of being viscous or liquid-like does not work for these techniques, it sags, it is gone.

In fact, we can hardly measure window glass, which anneals at room temperature, because the dissipation is so high that we get Q's of 10 or 20 on many samples; the resonances are smeared and it is hard to look at.

[Transparency]

We are going to talk a little bit about the wave equation in solids, and the beginning point, of course, is the standard one, where we have a solid elastic constant. Unlike the pressure or the compressibility of a liquid, we will have to be very careful about boundary conditions.

One engineering unit is Young's modulus, and that is a stiffness constant whereby if I stretch something, I apply stress on the end faces but I leave this surface free, so when I stretch it, it shrinks in diameter and that means that that modulus has both compression and shear, and we will get to that in a little bit.

I can set up, for example, a set of equations for something that has density in Young's modulus, but I might have to set up other equations as well, and I have to deal with shear.

[Transparency]

Let's talk a little bit about shear for a second. Shear is a volume-preserving distortion of a solid. Look at the little parallelogram here. It is pretty clear that if I took this square of material labeled A and I just shear it, I make this parallelogram.

But if I took material C and I sheared it this way, I also make a parallelogram whose shape is identical, it just happens to be oriented slightly differently. In fact, if I take this plate and I push it in at those two corners and stretch it at those two, I also make a parallelogram.

All of these sets of forces produce a shear, so there is a little bit of a choice as to the way I can define a shear and a solid, and I will pick one that is symmetrized in a little bit. That symmetrized shear looks something like this, where I claim that I have a shear strain, epsilon 1,2, that is, I have a displacement in the 2 direction along the 1 axis. If I take U to be a particle position, du1 dx2 +du2 dx1 and a put a half in front of it, now I have something nice and symmetric.

I have to pick a whole bunch of these shear pieces in, shears in different directions with different moduli, as well as the ordinary compressional stiffness pieces to write the energy of the solid. That makes for enormous complications.

[Transparency]

In fact, the full equations of motion for a solid turn out to have a bunch of fourth-rank tensors in them; that just means I have something with four indices on it. It is pretty clear how those are going to work. I am going to connect things like a strain, which is a second-rank tensor, for example. I have displaced the solid in the 1 direction with a force along the 1 direction. That might be a stress sigma 1,1.

Then I get several strains. If I squeeze it this way, I get a strainm an epsilon, in the one direction along the 1 direction, but I also get strains that vary in the 1 direction along the 2 direction as it bulges out, so I have a huge number of things. I have to connect stresses with two indices to strains with two indices. I need things with four indices, so the elastic stiffness tensor for the solid is a fourth-rank tensor.

All it means is I am connecting all these weird strains together in some nice linear equation, and we are going to look for solutions that e^{ikx} -omega T for this, except now we are in 3D with all sorts of weird indices floating around, so that all of the waves in the solid are described by this.

Most of the waves are neither shear nor compressional; they have K vectors in some direction with momentum transfer in another and displacements not parallel or perpendicular to either of those directions. So the problem is nicely complicated, but it is very rich in information.

[Transparency]

It is also sufficiently complicated and it yields to the usual physical properties of tensors that anything really nicely physical like this has to be symmetric, so there are not so many independent elastic constants in a solid; in fact, there are only 21.

That means you can kind of reduce the number of indices in that tensor, and this is a description of how the indices are collapsed, for your reference, so if you see something later that is C_{44}, C_{44} means it is really lambda 2,3,2,3, which is just a way of decoding it.

Another thing that is nice about many solids is that, for example, almost everything that we make stuff out of has a symmetry in which the crystallographic axes are all at right angles to each other. What that means is that the elastic stiffness tensor, when I relate stress to strain, has zeros in all of these regions here and that this region down here is pure shears that are volume-preserving, and this region up here includes things that look more compressional-like.

[Transparency]

This sort of behavior has subtle modifications. For example, if I have a single crystal with a little atom and a big atom -- I have drawn it that way here just to be simple -- and all of these little barbells are on a cubic lattice but they are oriented at some angle to it, when I squeeze this stuff, it rotates.

In those cases, when I squeeze it I get strains this way, I get strains this way, I get strains at odd angles, and I have a complete mess, in which case I need every single entry in that elastic modulus tensor to describe it, but I also get to extract tons of physics from it, if I can do it.

Here is the tensor for a cubic-or-higher symmetry. As you keep reducing the crystallographic symmetry, you get fewer and fewer elastic moduli.

[Transparency]

So when you hit cubic, I have got lots of zeros in the problem, so you can see, for example, that a shear strain in the x-direction connects only to a shear stress in the x-direction, because this is a diagonal piece of the matrix.

You can also see that I keep all the volume non-preserving stuff in one location, so it becomes very easy to deal with, and there are only three moduli, one shear modulus, one longitudinal wave modulus, and something that describes how much it bulges when you squeeze it, so for a cubic I am down to only three moduli.

The way you use these sorts of things, if you want to do a computation -- for example, let's say you want to compute the bulk modulus of an anisotropically elastic solid -- what I do is I might take the stresses in the x, y, and z directions and set them to one, and then I solve these equations for the strains, compute the volume change, and I have got the bulk modulus.

For example, a simple one is to compute the engineering modulus, Young's modulus, the one where I just stretch it one direction and leave everything free. There I have a stress in the one direction -- here -- and every other stress is zero. I have no shear and I have no stress on the other faces, so the shear strains are zero, I can solve these three equations here for the three unknowns, and, boom, out comes Young's modulus from the elastic moduli.

It is very easy to deal with and I guess there are MatLab programs and all that to play with these.

[Transparency]

The key problem and the problem with dealing with resonances is the bulging. This makes the computation of resonances nasty, because if I want to measure the elastic properties of something, I kind of need to have resonances that access all of them.

In order for that to happen, an object has to be small, compact, aspect ratios near unity. If I have aspect ratios near unity, when I start it resonating it bulges and then necks down -- I am just showing a cylindrical resonator here. There is a substantial amount of energy in the bulging motion, which means that the frequencies now are no longer simply related.

For example, if I have a short bar, the first half-wavelength resonance in it does not have a frequency that is half of that of one with two half-wavelength resonances in it, so the progression is no longer simple; every single mode needs to be computed.

That is what we are going to talk about in the next section, the computation and actual measurements. I think I am going to take a break here for 10 minutes, and then we will start on the next part.

DR. MIGLIORI: This next section is essentially required for resonant ultrasound. It has to do with the computational problem in some detail. It is interesting and it is important to see it, because it will make you aware of how complicated a problem like this can get and, also, how it is actually possible to wade your way through to a really nice solution.

[Transparency]

The computational problem itself is a massive computer problem. In fact, when Bill Visscher and I first started working this problem back in, oh, it must have been in 1987-1988, we had a Cray XMP computer and we set the program up in Fortran, which works fine and I still like it, in spite of the fact that it is out of fashion.

It was written in Fortran fairly carefully and I will show you what was written in a minute. The Cray at that time was bought by the nuclear bomb designers to make bigger and better bombs, but every once in a while they would let us use it.

We could run this code and it would take about a second per iteration to do this computation on what was then the most drool-inspiring computer in the universe. At the same time, we ran the thing on an 8-MHz IBM P.C.AT computer, with a 16-bit nice Fortran compiler on it. We thought they were the greatest things since sliced bread -- actually, I never understood why sliced bread was great. It would take 12 hours per iteration.

I now have this 600 Pentium III laptop computer with a nice 32-bit compiler on it and there you go, 1.2 seconds per iteration, same code, with minor, minor variations but, really, no different. That is my computation of how fast a modern P.C. is.

I am going to put off the motivation for the use of resonances and why we want to study elastic moduli until after I get through this computational section, because this starts to get to be really interesting and a lot of fun -- honest.

[Transparency]

We are measuring resonances and right now I will just blow off the complication of what the frequency is. We are going to just pick something close to <u>omega</u>$_0$ and hope we have a Q high enough not to intrude on the data.

By the way, you guys remember that there are lots of ways of measuring the Q, but you have to be careful to note whether you are looking at power or displacement, because there are factors of two floating around, and sometimes we will use the Q or the dissipation to extract physics.

The basic route to solving this problem is not finite-element computations. We could do finite element, but I am doing a measurement that is good to a part in 10^6 quite often. You cannot, even on a modern -- well, if the Los Alamos Blue Mountain machine actually worked, you still could not do a finite-element computation as well as you would like, that is, a part in 10^6 errors require meshes that have 10^6 elements in each linear dimension, and the computational intensity of the problem goes like a cube of the length.

But I can write down the Lagrangian for the solid and if the solid is homogeneous and it is continuous, it does not have to be isotropic, something really cool happens. So we begin with a Lagrangian, where we have a kinetic energy density and a potential energy density (notice that the elastic constants are now coming in), and I let the Lagrangian vary and look for a stationary point.

[Transparency]

I can then use Green's theorem to convert this volume problem to a surface problem, so now I have a computational problem in which all I have to do is get a bunch of functions on the surface of the object that I am measuring to show zero traction, that is, free surface, and obey this Lagrangian. That now becomes a much, much simpler computational problem.

Well, I do not need finite elements -- it is not simple by any means, but it becomes a very nice computation that you can do, and you can do it pretty rapidly. I am going to need to generate a complete set of functions to then adequately approximate a stationary Lagrangian on the surface of the sample, and the functions do not necessarily have to be orthogonal, which makes it really easy, because you can choose the set of functions that you use to approximate the vibrations of the solid to match the geometry of a solid, for example, spheres, cylinders, rectangular parallelopipeds, whatever you like.

I need a set of functions and I wind up, really, with a pretty straightforward set of equations to solve on the computer, a simple matrix eigenvalue problem, and that is not so bad to do.

[Transparency]

We have used trig functions for the basis set, cosines and sines, Legendre polynomials.

[Transparency]

And even simply powers of x, y, and z, depending on the geometry and the mood we were in. With solid-state systems that have orthorhombic, that is, all the angles are right angles or higher, symmetry, meaning orthorhombic, cubic, isotropic, the whole problem reduces to a set of block diagonal matrices, and you can do it very fast.

[Transparency]

So we are solving that eigenvalue problem that we extracted by finding the stationary points of the Lagrangian. That algorithm, as well as our Fortran computer program, are available to any U.S. citizen who asks, and I think we can ship them to non-U.S. citizens, but it is going to require about 80 pounds of paperwork (it did not used to). This is a big problem with this craziness.

What I have just shown you -- you have to think about this for a second -- is a route to computing the normal modes of vibration of a homogeneous solid of well-defined shape with zero dissipation, but that is not what I want.

I want the stiffness constants. So there is now an inverse problem. What I am really after is I want to take all the measured normal modes and compute the elastic moduli to produce this. This turns out to be a really snotty problem. There have been a lot of approaches to solving it. One that we use now is a modified, kind of steepest descent, minimization with, unfortunately, a fair level of operator expertise required when you are looking at 200-micron samples and you have absolutely no idea what their elastic moduli are. I will show you that in a second.

So that is the route through the computational problem.

[Transparency]

Now I will go over some other details about the experimental situation in general. I want to try to compare -- I am going back to hardware for a minute -- the signal-to-noise ratios one might expect in doing a pulse-echo measurement, you know, ping, time the pulse, measure the elastic constants to a resonance measurement. I am going to go through pieces of the measurement here on this slide and estimate signal to noise.

In a pulse-echo system I want to measure elastic constants. The sound speeds are the square root of an elastic constant over a density, so if I measure the speed of sound and the density I can get the elastic constants pretty quickly. One way of doing it is to take a big block of material, polish the faces flat, glue a transducer on it, ping it, measure the time of flight, measure the length, I have got the speed of sound, I have the density, I have the elastic stiffness.

Then you take the transducer and put it over here and you measure the next elastic constant, then here, then you glue on a shear transducer and you measure three shears, and then you cut the sample at 45° and you put them on, and pretty soon you have the whole elastic tensor. That is one way of doing it.

That technique requires that I ping the sample and then wait and then time the response, so that is a pulse-echo measurement. When I do it, I might put 160 V into a 50-ohm transducer for 1 nsec to generate a sound pulse. I could probably push pulse repetition waves up for small samples to 10 kHz, so the average power is on the order of a few milliwatts for this.

In a resonance system I put 5 V into a high-impedance transducer, so I have the same average power in both cases; I am trying to be reasonably generous. Remember, an order-of-magnitude difference here probably does not mean much when we are all done.

The measurement bandwidth: in a pulse-echo system, nanosecond pulse, 10^{-9} seconds, hit the sample. In order to have lots of accuracy, I had better be able to measure the time that I receive that pulse to 10^{9} seconds. That means my electronics has to have a bandwidth of 10^{9} cycles per second, a gigahertz. The noise bandwidth, my electronics is wide open, I am looking at a whole gigahertz of Tom's noise there. So I have a problem.

Resonance measurement: Actually, I am being very generous here to the pulse-echo guys, because if I were making a resonance measurement and I have, say, a Q of 10^{4}, which is quite common (that means a megahertz resonance is only 100 Hz wide), I need to look at only a few resonances, maybe, say, 10 of them, for an isotropic material and they are each 100 Hz wide, so I really need to measure only about a kilohertz of the universe.

It is very easy to get a small number there. The noise bandwidth for the resonance system -- I mean, I have to measure, maybe, over a whole kilohertz, but what is even better is I can take my time, I can sit there and measure at a frequency, step the system, measure again, so I can average as much as I like. I can actually make the noise bandwidth very, very small, but we will leave it at a kilohertz for the purposes of comparison here.

Electronic noise: There is no good reason why these should be different, so I have not made them different and, in fact, although I do not quite hit 1 nV prHz on the system I built, I do do 3 nV pr Hz. There are good reasons for that. Those are about the same.

Then I have the detection duty cycle. How long is it, how much of the time am I actually processing useful information? In a resonance system it is 100% of the time. In the pulse-echo system there are problems. You are really only acquiring information during -- well, I do not know if I want to go into real depth here, but think of the following problem.

When I do not have a pulse present, look, I have pinged the system, boink, boink, when there is no pulse present, it is just as important as if there is a pulse present, so you cannot really say that your duty cycle is a nanosecond at a 10-kHz rate, but it is still not very high, so I have been somewhat generous here and put in .01.

Now I can compute the overall voltage signal-to-noise ratio by combining all of these factors. You see it is that not different; resonances are better by a couple of orders of magnitude but it is not seven or six orders of magnitude. So resonances are nice and they help; if you are in a noisy environment you would probably want to use them, but they are not maybe as wildly good as you might have expected. I did this once before and got the wrong answer, so I needed to put a better answer up this time.

[Transparency]

Maybe measuring resonance is 100 times better than measuring a pulse-echo response in a system if you really look at it carefully -- well, sometimes a factor of 100 is nice.

DR. GARRETT: How do you factor in all the time it takes to unglue it and reglue the transducers?

DR. MIGLIORI: Thank you, Steve. Not only that, I have to make bond corrections, which means I have to think and I have to write something in my lab notebook instead of reading it from the computer screen. If I think, I could get fired. (laughter)

No, I am paid only to fill out security plans and standard operating procedures. If I am caught doing any work, it had better be after hours. (laughter)

That was a rough description of signal to noise. There is a whole bunch of little details on getting resonant ultrasound to work. For example, I probably want elastic constants to a 10th of a percent when I am done, which means I have to take little 1-mm samples and get their faces parallel and perpendicular to a 10th of a percent, which is a micron.

We could not figure out how to spend the government's money, even though we had plenty of it, to buy something that would do that, so we came up with a simple scheme. This is a Pyrex plate. These are gauge blocks -- have you ever seen these in the shop? -- you buy four sets of gauge blocks. You put the gauge blocks down on this plate and then you take your sample and you catch it in the middle -- you see I can push this block and this block together -- put a magnet under there and a little heat-sensitive glue and I can push those four things in to hold the sample.

Then I can polish it on an ordinary piece of sandpaper, flip it over, and polish it again. Now I have two parallel faces. The problem is getting perpendicular. So now I take those two, bring them up, push in with the gauge blocks, and I get a perpendicular face. So pretty quickly, using $20.00 worth of stuff, we can do better than the the several-thousand-dollar polishing rigs that you may occasionally see for sale.

Solving this problem got us from a nonworking experiment to a 10th-of-percent experiment, because if you make little errors in parallelism, these resonances start to cross each other, that is, the thing you thought was the 18th resonance is now in the 19th position. You completely misinterpret the spectrum. You try to fit it, you get big errors, you are dead meat.

The technique actually will not do much worse than about half a percent. If it goes beyond half a percent, it instantly and nonlinearly scrambles and you get nonsense.

MR. APOSTOLOU: How long does it take you to do one sample?

DR. MIGLIORI: How long does it take me? How long does it take a new postdoc? How long does it take an experienced postdoc? Anywhere from an hour to a week-and-a-half.

For example, it might take me a week-and-a-half, because I can work on it only about 30 seconds at a time. For example, we did a measurement on Europium hexaboride, single crystal. The crystal was 200-microns thick, about 800 microns long, and maybe the other dimension was 300 microns. That took us about a week-and-a-half.

First, we had to x-ray-orient it to find the crystallographic axes. We had to then mount it, take that goniometer and polish one face of the crystal. Then we had to remove it without cracking it and proceed by making shims that were the right thickness, which we had to actually surface grind. But an ordinary piece of steel might take an hour or less; very easy.

[Transparency]

This is a typical, meaning the only, commercial room-temperature resonant ultrasound stage. There is a hinge here, so I can pick this up with my fingers, use a tweezers to put the sample in there, and I am done. That was the transducer-mounting problem.

You talk to the pulse-echo guys and that is why you see a lot of dust on their pulse-echo equipment after they get to play with one of these.

[Transparency]

The transducers are not quite trivial and, in fact, we had to spend a lot of time making sure that those transducers -- remember, we have done a careful job in electronics design. We got lots of signal to noise. That electronics package will pick up anything those transducers produce.

One of the things transducers like to produce is whatever happened to the sound that goes from this transducer through the base, up around here, through here, and back to here? It is very easy to get lots of signal when you do not have a sample.

Also, if you are a commercial designer of electronics and you decide to build a resonant ultrasound system and you did not pay too much attention, you can also get signals when there are no samples present. One of the competing manufacturers did this by taking this 2 nV prHz receiver and sticking it inside the computer.

We need to isolate the transducers themselves from their surroundings and what we used was a glass, a glass that is really a nice glass at low temperatures, namely, Kapton. It is a polyamide film, it is an orientational glass, its ultrasonic attenuation increases as you get it colder. Glasses tend to do that.

For example, in fused silica the ultrasonic attenuation goes up by a factor of 200 when you cool from room temperature down to 40 Kelvin. This has to do with the fact that as you cool, kt starts to approach the energy required to reorient molecules in the glass and when kt equals that, they go crazy. You get big attenuation peaks.

We used glass, namely, Kapton, and we glued small single-crystal lithium niobate transducers. These are piezoelectric, not ferroelectric, there are no domains, the hysteresis is zero, for all intents and purposes, and they have no dissipation, so they are strictly piezoelectric. You squeeze them, single crystal, they make electricity.

That kept dissipation there down and then we messed that up by putting these plastic leads on it. Plastic leads turned out to be really important, because any metal conductor we use would carry sound right out the lead. Jay Maynard has been using polyvanilidine fluoride, which solves this same problem by using a plastic transducer. There are a lot of good ways to do it. This is the one we like and it has very successful.

Then we have a backload and mostly we have using single-crystal diamond for the backload, because its sound speed is so high that when I bond it to this mess -- the resonant frequency of this is very high and it does not really intrude very well. The little bit of damping here keeps the Q's, when we hit the first resonance, is low enough that we can usually see the samples and select them out, so there is some care in this design.

[Transparency]

We do a lot of low-temperature measurements. I have a 10-pF transducer there. That transducer is a millimeter-and-a-half in diameter. I have to get a signal from a 10-pF transducer up to 2 m of cable into the room-temperature electronics, another not-so-nice problem.

Here is how this one goes. I put together an op-amp circuit and I choose things so that the noise voltage from this resistor at the frequencies at which I measure is completely short-circuited by the capacitive reactants. This, then, gives me DC stability for this op-amp so that it does not run away. If I moved this resistor, it would integrate its output and just run to the rails.

I see some blank looks? Okay, I do not think I am going to give an op-amp course right now, but suffice it to say, when you design something with op-amps like this and you have a capacitor in the feedback circuit, it has no DC path. Therefore, the op-amp's input errors, for example, it might produce 25 μV of input offset voltage, which means it is a voltage that is always present and it then attempts to amplify that, although I have negative feedback meaning I am taking the output and making it talk to the inverting input of the op-amp.

If I have no DC path, I cannot correct for that offset and it will just keep trying to remove the offset until it gets up to the power-supply voltage and quits, so you have to have a DC path, but I am trying to measure transducers with a few picofarads of impedance, so I cannot have more than 10's of picofarads here.

If I have 10's of picofarads here and I am interested in measurements at frequencies of hundreds of kilohertz, this thing has to be 10's of megohms. What is the noise voltage off a 10-

megohm resistor? At order of 30 nV prHz, my bandwidths might be lots at these frequencies of kilohertz. That resistor produces tons of noise.

As long as the frequency at which I make the measurement is such that the impedance of this capacitor is low compared to the resistor, I will not have a problem. What happens is at low frequencies I do see this and all my electronic sees is this enormous low-frequency tail to the noise, but as soon as I get into the 10's of kilohertz range, this thing shunts the noise voltage generated by this resistor and it starts to look as if I have a pure capacitance in there.

The key is that this op-amp has to first order an infinite gain. This terminal is connected to the ground, zero voltage. The op-amp is going to do everything it can to keep this terminal always at zero voltage, but the transducer is generating charge as it flexes, say, at hundreds of kilohertz or megahertz frequencies.

The transducer pumps charge down this cable here trying to raise the voltage of the center conductor, but the op-amp says no way, it sucks the charge off this way, keeping this at zero voltage. You would say to yourself, gee, it does not do anything but, in fact, it does.

When it pumps charge this way, this stays at zero, this thing has to start following the charge, so the voltage out here is proportional to the charge transferred in here, but there is a really cool thing that happens. The op-amp is keeping this at zero voltage all the time. The outer part of the coaxial cable is at zero voltage. So what is the difference in voltage between the center and the outside? Zero.

As long as the op-amp is fast enough to keep this process going, the cable capacitance does not short the transducer out, so I can get my 10-pF transducer, I can get the full signal right up to room temperature. There are lots of other tricks like this in the electronics and some of them I am going to show you in the last part of the talk.

[Transparency]

Finally, there are bits and pieces of black art floating around. This is a measurement on some material (I forget what it is). I made a measurement. I have measured these resonant frequencies here and I attempted to do a fit to elastic moduli and I find there are a couple of resonances that just have these great big fat errors. Oh well, what did I do wrong?

If you look at it very carefully, you can see that this resonance really corresponds pretty well to the next one down, and this one here, and so on. What happened is I missed a resonance, something went wrong, there was a node on the transducer for a particular mode; I swept right through it and never saw it.

But the pattern here tells me what to do. What I do is I tell the computer, well, I missed a resonance, so we will just put in a zero there and tell it to ignore that in the computation. Now all of a sudden I get everything lining up and my errors go from this to this on the same data set just by my guessing that at that point I missed a resonance.

There are several pages in the book on how to make that guess better, but this is an example of the primitiveness of the computational procedure, at least the way I do it. In the end, we win, and we will go back to those beryllium data again.

[Transparency]

I want you now to look pretty carefully at this column -- right here. These are percent -- let me say it, again -- percent errors between measured frequencies and the computer-generated fit by juggling elastic moduli -- .09%, .02%, .01% (there is a bad one at .02%). That is what this thing does.

[Transparency]

So we can use it to do some physics. This was a measurement of a colossal magnetoresistance material and it was quite interesting, because this material was rhombohedral. Can anybody tell me what a rhombohedral crystal structure looks like in this room?

DR. GARRETT: The same as it does outside.

DR. MIGLIORI: Excellent. (laughter)

[Transparency]

Rhombohedral: I take a cube and I take it along the body diagonal and I stretch it, so rhombohedral is a cube with the body diagonal stretched a little bit. The crystallographers show us the crystal structure of this system as rhombohedral.

You look at it and it is a mess, because they always draw it -- in the crystallography books they will draw the crystal structure in such a way that its "rhombohedralness" jumps out at you, but what they do not tell you is that it is only maybe 4/10% distorted from cubic.

I do not have a code to do rhombohedral very easily and I probably need to measure 30 resonances to get rhombohedral elastic constants if I wanted to use that messy code. So we tried to fit a cubic. It fit perfectly. We got sub-10th% of a fit of cubic elastic constants for the system.

What we found, which was very interesting, is that the single crystal had stacked up with the rhombohedral axes located randomly, just like the domains in your ferroelectric, but the distortion is so slight that even over macroscopic distances the net displacement of an atom from perfect cubic symmetry was very small, a fraction of a unit cell.

That turned out to be really important, because it showed that there was an instability to that lattice that helped to understand a lot of the properties. For example, application of a 5-T magnetic field with shifting elastic constants 15%. What it was really doing was realigning some of these little rhombohedral distortions, we think.

A measurement of the full elastic tensor with resonant ultrasound told us that this rhombohedral system looked cubic and that, therefore, it had to have a fine-scale, very weak domain structure with a very small energy difference between these and, therefore, there was a mechanism for clicks in elastic constants as you apply magnetic fields.

To do that with pulse echo, you would not have found out, because either -- pulse echo always gives you an answer, I always get a speed of sound. I do not know if the crystal was rhombohedral or cubic, but resonant ultrasound will say if this were rhombohedral, and significantly so, I would not be able to fit cubic elastic constants to it.

I might have missed that whole point and I did get the entire elastic tensor, so I noticed that in certain directions in the magnetic field there were no effects on elastic constants and from that we can kind of eke out what the physics was.

[Transparency]

This is a martinsitic transition in a gold-zinc single crystal. These martinsitic transitions are sort of an ordinary phase transition, except what happens is some part of the crystal starts to go from one crystal structure to another. It generates a strain field, which then hangs up the transition nearby until the temperature drops a little farther, at which point this one goes and now builds a bigger strain field and forces the transition to hang up a little more.

You lower the temperature, and so the whole transition takes a while to do and kind of makes a mess ultrasonically of it, but here is just an example of a bunch of resonances measured on this crystal as a function of temperature and then it goes martinsitic and they go through the floor; you get large shifts. I am just showing you bits and pieces of measurements.

[Transparency]

Here, again, was the superconducting transition in niobium. That is interesting, because -- was it Peppard? maybe -- in 1955 did an analysis of what happens to the elastic constants at a superconducting phase transition.

Superconductivity, the old-style one, was a perfectly good scaler quantity. There were effects in the specific heat, and so on, but there was no order parameter that had a direction to it in a simple type I superconductor and, therefore, there was no expectation that if I sheared the superconductor this way, or this way, that I would have anything different happen, it should be the same, this shear and this shear, all the physics should have been the same.

From that you can bootstrap yourself into realizing that the elastic moduli cannot have a discontinuity at that phase transition, they can have only a break in slope, and there it is, a break in slope of the shear modulus, and about the right size, and so on.

[Transparency]

Another interesting thing that we did was to attack the electronic subsystem in a narrow-gap semiconductor. The compressional stiffness of a solid also includes compressing the electrons as well as the atoms and the vibrations.

What we did here was to assume that this was a narrow-gap semiconductor, so it had electronic densities of states just on either side of a Fermi energy, if you will, that were very narrow and peaked. From that you can compute the elastic constants that you would expect.

When we fit that, we found that there were very narrow peaks in the electronic density of states along the gap. We could fit it very well to a function that generated this and we were able to predict quantitatively the electronic density of states in this system. It disagreed completely with microscopic photoemission measurements and we were into a little controversy.

Those measurements were then redone by someone at Stanford and they found, by going at very high resolution with their photoemission spectrometer, these two peaks in the density of states that we had predicted based on a benchtop classical physics measurement.

[Transparency]

That was kind of the nuts-and-bolts and a little bit of what we measured and at this point I am going to start on the real motivation for making resonant ultrasound measurements and we will finish up with electronics.

[Transparency]

I am going to start with a review of statistical mechanics, my favorite thing. This stuff is going to be applied to plutonium. I am going to convince you that the study of acoustics is going to tell us how to understand this mess.

These are the crystal structures of plutonium metal between room temperature and about 600 Kelvin. It starts out at room temperature in a monoclinic, which is a kind of sheared strange crystal structure with 16 atoms per unit cell required to describe this element.

You warm it up to about 100 C. and it goes to a body-centered monoclinic crystal structure with 34 atoms per unit cell. You warm it up some more and it goes orthorhombic, which has now everything at least at right angles but none of the axes are the same.

You warm it up some more and it goes face-centered cubic. Now we are starting to get to crystal structures that you can actually deal with. Finally, you warm it up again and it becomes body-centered cubic just before it melts.

All right, that is about as many phases as anything but cerium has and it is a puzzle, it is a very big puzzle that is not understood. Most metals will have two phases, maybe fcc and bcc and very few metals have anything lower in symmetry than hexagonal; certainly monoclinic is a mess.

There are other very weird things going on with plutonium. For example, I am up at a moderate temperature and it becomes face-centered cubic. Now I start to warm it and the volume contracts on warming. Throughout the entire existence of this phase the volume thermal expansion coefficient is negative.

Then it makes a phase transition to body-centered cubic. There is a big 3.5% volume drop when it heads for body-centered cubic. Now the thermal expansion coefficient turns positive, a nice small positive value. When it melts, the volume drops another 3%.

Studying plutonium is an intellectual puzzle that we have been denied, because of the nuclear weapons problem, until recently, when this metal has been made available to us to do basic research. This is a fantastic problem. It has every piece of weird metallurgy that you might find in six or 10 or 20 other systems, all combined in one system.

Except for the fact that you have all those guys with machine guns watching it, it is a great metal to study.

[Transparency]

I am going to show you that the single most important tool for studying plutonium is acoustics, period, no question. To do that, I need to lead you through statistical mechanics.

It is easy to forget what is going on here, because people talk about things like "free energy." I have absolutely no idea why they came up with that phrase. I know exactly what it is, I know exactly what it means, but I do not understand where those words come from and I think it has to do with the Kramers-Kronig problem. You know, when you get your Ph.D. they take you into a little room and they give you the secret decoder list for all these physics phenomena, like "this simple concept is free energy," "that simple one is Kramers-Kronig," and you are not allowed to tell anybody else.

But we are going to get through this, so I think you will at least have a gut-level intuitive understanding of free energy, and I need to get there so I can get to ultrasound. To get there, I have to start with something simple and what it is going to have to do with are numbers, numbers of objects.

We are going to look at some really simple system like flipping coins, and Jay Maynard is going to let me get through this without throwing oranges at me. We are going to look at something really simple, like a few coins that we are going to flip. Then we are going to imagine that we can extrapolate this to 10 of the 22nd objects. What you will find is that approximations that are crude and very weak for 8 coins become so close to exact for 10^{22} objects that we are very happy with the problem.

[Transparency]

We will start doing a little statistics, first, and kind of easy statistics. We are going to flip coins. I have 8 coins and I flip them and I look at what I have on the table. Sometimes I get one head, sometimes I get zero heads -- 1, 2 -- so here is the number of heads. I could get 0 to 8 heads.

Then I look at how many ways I could get zero heads. There is only one. All the other coins have to be tails, but there are 8 ways I could get one head to show. The first one could be head and all the rest tails or the second one could be a head and all the rest tails, and so on. So there 8 different ways of getting one head and there are 70 different ways of getting 4 heads.

If I have no reason to believe which particular throw will produce what, I have to assume that I sample all the possible arrangements equally. When I do that, I find that, on average, 70 of my throws will be equal numbers of heads and tails and then 56 out of every 256 throws will be 5 and 3, and so on. It is a very simple calculation, you guys can all do it pretty easily.

I get a distribution that tells me what all of the states of this system are and how likely it is I will get them because I see how many possible ways there are of getting each state of a given parameter, the parameter being the number of heads.

Let's see, 70 out of 256 is 40% percent of the time you get half and half and then it decreases. The interesting thing to note about this system that has 8 objects and 256 possible states is that the least probable state has a probability of 1/256 of being observed, and the most probable state has a probability of 70/256 of being observed. That is almost two orders of magnitude difference; that is not very much.

I do not know how carefully I am going to show this to you, but as the numbers get much, much bigger, the sharpness of this distribution function gets much sharper as well. For example, if I have 10^{22} coins, there is still only one way of getting all heads, but there is of order two to the 10^{22} ways of getting 50-50.

So the ratio of the peak height to the wing is very much, much greater, and I am starting to get numbers whose exponents have exponents in them. When I get to big numbers, this distribution becomes extremely sharply peaked, and stuff like that. You remember all that from stat mech.

[Transparency]

There is a ton of incredibly clever math that goes on. For example, the number of states with N heads, if N is the total number of coins, and n is the number of heads, then the number that have n heads is this thing here. Let's say I want 3 heads, so I have 8 choices for the first head, 7 for the second, 6 for the third, but there is 3 x 2 x 1 ways of arranging those. The answer should be 8 x 7 x 6/3 x 2 x 1.

Well, 8 x 7 x 6 is this piece here, and 3 x 2 x 1 is that piece, so that gives me the number of states. It also turns out that this is the pre-factor in expanding A+B to the end. I can see right away that if A=B=1, 1+1 is 2, 2 to the n, t is 256, the total number of states we have and that is just the sum of all these factors. I always like these things. I do not know how they connect up so nicely, but that is wonderful.

There are other fancy math tricks in stat mech that are nice, but when we start to push these to large numbers, we are getting numbers like this. One of the interesting things to ask yourselves is, well, I now know how to calculate this probability distribution and with a few approximations I can do it for 10^{22} objects.

If the most probable state has a probability P, what state has a probability P/2? Forget what the probability is, just ask how far do I go on the distributions before is drops in half. It turns out that if you work this problem you get a square root into it.

In 10^{22} flips it would be reasonably likely to get 10^{11}th more heads than tails. That seems like a big number, but 10^{11} out of 10^{22} is only a part in 10^{11}th, so the errors in my experiment when I do 10^{22} coins are parts in 10^{11}th. Even at Los Alamos we have trouble measuring physical quantities to that accuracy.

Big systems, the effects on large numbers: It is going to clean this problem up. The statistics produce things that we can count on to be essentially exact. First rule: All the numbers are going to be large. We will make it so. Even with nanomaterials we can easily get 10^6 or 10^7-10^8 particles in little tiny masses and springs fabricated on silicon chips.

It turns out the other half of the errors, instead of being square root are of order log, so the log of 10^{22} is 51, which is pretty small compared to 10^{22}, so once we hit these size numbers, we can do almost anything we want, as long as we introduce operations that will produce errors like square roots or logarithms.

[Transparency]

Next rule: Every system has lots of possible states, the coins on the table. For the coins there is no reason to expect that you would get any one particular of those 256 states over another. Of those 256 states, 70 of them have one property in common but they are still individual states, so every one of those states is equally likely.

Physical system explores, as time progresses, all of the states that have an energy very near to the energy that it has. I am using "very near," because I am going to get tied up in the difference between classical and quantum statistical mechanics and I cannot, for the life of me, understand classical statistical mechanics. The quantum stuff is infinitely easier, because I can count everything (I will show you that).

Here I have lots of different possible states and we are going to assume that the system explores all of the ones that have equal energy. There are circumstances where this so-called ergodicity hypothesis and those have to do with physical constraints, like sometimes the particles are in a box, so you cannot have them explore a state in which they are in a spherical volume because of the constraint, simple things like that. Then there are more complicated ones.

[Transparency]

Then we can do the classic state mech computation: I have the particles in a box, what is the probability that any atom of gas is in the left half of the box? It is a half. What is the probability that half of the 10^{22} of them are in the left half? It is a half to the 10^{22}. What is the probability in the life of the universe that I will see that happen?

The probability that any one atom of gas is in the left half is a half. The probability that all of them are in the left half is 2 to the -10^{22}. The number of times that we observe this happen in the entire life of the universe? Let's say that we do the experiment at a rate of 10^{11}th times per second, relating to the collisions between gas molecules.

So we try the experiment 10^{11}th times per second, which means that in the life of the universe we will try it 2 to the 402 times. It looks like a big number, yes? The probability is 2 to -10^{22} that they will all be in the left half, so in the entire life of the universe we do not even come close to seeing this happen.

It turns out, though, if you start to ask how many boxes are in the entire universe, so that the number of times you try it on each box is 10^{11} but there is something of order 10^{60}th boxes, there may be a chance. But you cannot say every once in a while you are going to see this box with everything in the left half; the probability is not different in the entire life of the universe than it is one second to any appreciable degree.

Okay, big numbers, nothing improbable is going to happen. That is really important.

Last rule is that the total number of states is a very strongly increasing function of the total energy and I can do this only with the quantum system, so let me show you the quantum system, because I do not know how to do it with the classical one.

[Transparency]

Here is the quantum system. It is three harmonic oscillators sitting there. Harmonic oscillators have equally spaced energy levels, so this is at ground state, first excited state, second excited state. The frequency of the oscillator tells me the energy separation, so as I add a quantum of energy, h bar $\underline{\text{omega}}_0$, I go from here to here. When I add a second h bar $\underline{\text{omega}}_0$, exactly the same energy, it goes to here, and so on. It always exhibits the same frequency.

[Transparency]

If I have two quanta of energy in my system of three harmonic oscillators, then there are six possible states. I can have two quanta in the first, zero in the second, zero in the third, and so on, six states. Now I put in three quanta of energy. It turn out I now have 10 states -- just count them up. You see how easy this is in quantum mechanics, I can count the states, I just know what they are.

If now I have a big system that is composed of groups of these three harmonic oscillators and I add energy to it, if I increase the energy by roughly 10%, then each subsystem has its energy increased by roughly 10%, on average, and, therefore, it has 10% more states than it did before that it can access.

If there are 10^{22} of these, then I take 10%, which is 1.1, to the 10^{22} and that tells me how many more states I have when I added a little energy to this. The number of configurations, the number of states, of systems, is a very strongly increasing function of energy.

[Transparency]

Okay, I am set.

Let's just do a little manipulation now. I have a large system whose total energy is E -- wait, let's stop and have a break and we will continue afterward.

DR. MIGLIORI: Actually, we have a chunk of copper here and we have drawn in Magic Marker a little line on it. Everything outside the line is part of the larger system. Everything inside the line is part of the smaller system. We noted before that as I increase the energy of a system, the number of possible configurations increases wildly and, correspondingly, if I decrease its energy, it drops rapidly.

I might ask, what is the probability that the smaller system has some energy E1, whatever that energy is. It is proportional to the number of configurations that the smaller system has with energy E1 times the number of configurations that the larger system has, if it has an energy E-E1 -- an absolutely trivial statement.

[Transparency]

But now I want to know what is the most probable energy E1 that is in the system. What I do is I try and find the maximum and the probability P, so I take the derivative of P with respect to the energy of the smaller system, set it equal to zero, and I get this curious piece here.

I notice that d log $\underline{\text{omega}}_1$ de_1 equals d log $\underline{\text{omega}}_2$ de-E2, which is the same as the energy of the smaller one. So I found a property that, if I put these two systems together, they have in common; that is, the probability distribution is very sharply peaked. I am going to see in the life of the universe only the most likely configurations, meaning I am going to see the group of configurations near this particular energy, because everything else is in the tails.

When that occurs, it is a particularly interesting state. It says that the fractional change in the probability of one of the small system, if I remove energy from it, is matched by the fractional change in the opposite direction of the probability of the larger system as I added to it, fractional change.

That says that these quantities are the same if I look at the system and let it equilibrate, wait for a long time. From that comes the definition of temperature. It is d log omega de, the inverse, with a k Boltzmann in front of it. That is the fundamental definition of temperature, not kinetic energy, not anything else. It is related to the derivative of the logarithm, the number of configurations with respect to energy.

Then we can define entropy as the logarithm of the number of configurations and that is pretty cool, too, because that is a way of making an extrinsic quantity of entropy. That is, two pounds of butter have twice as much entropy as one pound. I did that out of things that are being multiplied together. Remember, the probability is the number of configurations in the smaller system times the number in the larger, so I now figured out, using logarithms, how to make an additive quantity of things that are multiplied together. There is temperature and entropy.

[Transparency]

This is all very puzzling, because the net result of this is that the universe behaves in a pretty strange way. Let me give you a simple example. I have a warm glass of water. I cannot tell, because I am always looking at the most probable configuration, whether 10 minutes ago that was a warm glass of water or it had hot and cold water poured into it and they have mixed. I cannot tell what that system looked like 10 minutes ago.

But every single physical law that goes into describing that system is time-reversal invariant, so something went wrong somewhere. In fact, I have taken a quote from Landau and Lifshitz that really describes this, and you can read it for yourselves, but they and other people, myself included, believe that there is some fundamental thing very strange about that particular property that defines an arrow of time. It tells you that the entropy is always increasing, because whatever I do, that probability is always heading toward the maximum value as time progresses.

But let me use it for a second now that we have that. I will do a simple Taylor expansion. I have a system with an total energy E and a smaller system with energy e_i. I want to know how many configurations the larger system has as I juggle energy into and out of it, so I will Taylor-expand it, so log of omega of E-e_i is just log of a system as if it had the total amount of energy minus d log omega de_i. A simple manipulation tells me that the number of configurations with energy E-e_i, that is, I have taken from the bigger system, over the total number of configurations of the big system is the Boltzmann factor.

So now I have the Boltzmann factor out of this simple thing. To get actual probabilities from number of states, I simply have to normalize it, so now I just sum over all the possible Boltzmann factors, and I get the very famous partition function. So just this trivial manipulation with flipping coins and saying we are going to see the things that are most probable, and the large numbers, got us the Boltzmann factor and the partition function.

[Transparency]

The partition function describes kind of the number of configurations times the energy of every possible configuration that the system has; that is, instead of summing over every possible energy, I might group systems in different energies. For example, water has a certain energy per atom and ice has a certain energy per atom.

I will say, for the most part, every configuration of ice has the same energy per atom and every configuration of water has the same energy per atom. I am going to be off only by the square root of something like 10^{22} or the logarithm, so it is good enough.

I do that and I will rewrite that partition function as the number of systems that have energy e_i times the probability of finding each of those systems with energy e_i. Now I am going to further collapse the whole mess by saying that distribution is so sharply peaked that, in fact, I observed only one energy, E bar, and there are omega states with that energy and everything else is negligible, and those states have a probability E-e bar/kt.

From this I have got the free energy -- here it is. I just make this thing equal to the E-free energy/kt. If I take the logs, I find out that the -kt log z, the partition function, is the energy -T times the entropy, it is the free energy of the system. If I exponentiate the free energy, I get back to this piece here.

This is really important to look at, and let's look at it down here even more. The partition function kind of tells me how many states are water and how many are ice. At any temperature, I do not care what it is, let's say I am at 800 C., there is a probability that this system is existing as ice, as water, and as steam. How likely is it that it exists as ice?

The number of configurations at 800 Kelvin of ice is so small compared to the number of configurations of steam that even though the system is fluctuating into that over an infinite amount of time I very rarely see it, so it looks like steam.

As I get it very cold, the probability of steam is so low that even though steam has lots and lots of configurations, this weighting factor makes it so improbable that I do not see it, but there is a temperature at which even though liquid water has a higher energy per atom than ice, therefore, it has of order that energy difference to the 10^{22} more configurations, but every one of those configurations is less probable by $E\text{-}e_{water}/kt$ than for ice.

But there is a temperature at which the huge number of configurations of water matched by the fact that they are less probable equals the fewer number of configurations of ice, matched by the fact that they are more probable, and that is the phase transition temperature and that is what the free energy is good for, and that is it.

Our job with plutonium and everything else is to understand the bits and pieces that go into this. They are almost all the sound speeds, that is what controls it.

[Transparency]

Phonons: Unlike Tom's problem, we are going to have a finite number of modes when we are done with this. Solid material: I am going to model it in 1D just to make it easy. We will have a big long string of masses connected by springs, so the "..." means this thing going on forever. We solve that problem by setting up F=ma.

[Transparency]

Here it is. We have a discrete problem. We have discrete masses and discrete springs, so the displacement of the ith mass and the ith-1 mass, and so on, come into it. We find that the solutions are sine waves. The dispersion relation for this long chain of masses and springs is sinusoidal in the k vector, where k is 2π/<u>lambda</u>.

Sine wave is linear at very low frequencies. Those are the sound speeds, d <u>omega</u> dk of the speed of sound at low frequencies. At very high frequencies it flattens out.

[Transparency]

Here is a picture of a set of dispersion curves. Here is <u>omega</u> as a function of k. Every one of those is a normal mode of the system, every normal mode acts like a harmonic oscillator. All we have to do is figure out how to calculate the entropy of a simple harmonic oscillator. Add them up over every single vibrational mode in the solid and I know the entropy of the solid and the energy.

It turns out that in plutonium or steel or nickel, at 800 Kelvin, 95% of all the physics is the vibrational entropy of the solid. That is what is going to control the phase transitions, and I will show you how in a second. Let me hit this for a minute.

[Transparency]

We are going to dot our i's and cross our t's. The average quantum number for each normal mode can be found from the partition function by just noting that is the normalization factor, I want to find the expectation value of N. I put in the weighting factor, sum over all the modes. Because the modes are equally spaced, I can do the sum in closed form and if kt is much less than the frequency of oscillations of any of the modes, then approximately the average quantum number is kt/h bar omega. Modes oscillating at h bar omega ought to all have kt of energy, and Tom told us that this morning, therefore, the quantum number is about kt/h bar omega.

The energy is h bar omega times the quantum number plus the zero point energy, and that has to be equal to kt. That all works out.

[Transparency]

The key point is that TS, the piece in the free energy that determines the number of states -- not the probability but the number of states -- of any particular energy is kt times 1 plus the log of the quantum number. So now I take every single mode. Here is a mode, here is a mode, here is a mode. There is a discrete number of them because there is a discrete number of atoms in the solid.

For each of those modes I compute the frequency. I then compute the log and I use that to compute TS. In fact, this computation from the measured phonon dispersion curves that you get from ultrasound for an insulating solid produces specific heat numbers that are better than the measured value, more accurate.

But more importantly, it is going to tell us if we change the vibrational frequency spectrum, what is going to happen, and that is what we are after. Before I go on, this was computed for that simple mass-spring system. I can also take the mass-spring system and add second-nearest-neighbor springs. When I make those positive, I get a dispersion curve that looks like that. If I were to make it negative, it would go like this (which I forgot to put on this picture).

We made some attempt to use sound speeds in plutonium, very preliminary, and neutron scattering measurements, very preliminary, to generate dispersion curves for plutonium.

[Transparency]

Here you can see it. There is the second-nearest-neighbor piece with a positive spring coefficient. Here is the second nearest neighbor piece, with a negative spring coefficient. Those are different mode types, shear compressionals.

DR. GARRETT: Aren't you worried about scattering neutrons off plutonium? Isn't that kind of a bad thing to do?

DR. MIGLIORI: The problem is that you cannot do it on Pu-239. The absorption cross-section is so high, you cannot get enough neutrons to detonate a subcritical mass, but the absorption cross-section is infinite. You actually have to use Pu-242 for the neutron-scattering experiment.

It is very amusing, there are only a few pieces of 242 around -- anywhere. It generates the most amazing fights among scientists as to whose Pu-242 it is and which experiment they get to do, but I think I won about two weeks ago, so it is mine and I get to do what I want with it. That is a good point.

These indicate different directions in the solid, 45°, 90°, but we are getting a picture of it.

[Transparency]

What is really going on? Two phases of plutonium

-- look at this for a minute. This is face-centered cubic, except I have drawn it rotated 45°, so it looks like body-centered tetragonal. This is exactly identical to a face-centered cubic solid. That is one phase of plutonium.

Here is body-centered cubic. That is another phase of plutonium. Look what happens when I squeeze this. It becomes that. So try this experiment. I have a single-crystal plutonium. It is face-centered cubic and I am holding it in my gloved hands inside the glove box with the guys with the machine guns behind me.

There is no stress on this face, these faces, or these, or these. It is cubic. Now I squeeze it. I squeeze it so that the ratios of the legs change by the square root of two, making it body-centered cubic. Now it is body-centered cubic but I have no stress on this face, and no stress on this face. What is the stress on these faces? Zero by symmetry.

When I squish this stuff, I am compressing this metal and it goes "mmmm" and then "voompf" and when it reaches dcc it stops, there is no force. That is called a Bain route. That is a very soft route. It can fluctuate into that route, and I am starting to develop a theory of negative volume expansion, because it appears that the bcc phase has a smaller volume and it is fluctuating into that at higher temperatures, so the volume is decreasing.

DR. GARRETT: But there is an activation energy.

DR. MIGLIORI: There is an activation energy but I can actually, from the phonons, start to compute it.

[Transparency]

I just wanted to show you that route, because it is very interesting, but the other thing that is interesting is that in bcc plutonium I have not figured out how to draw it, but it turns out there is a bunch of ball bearings that are just touching each other, so they are kind of rolling like this.

If I look at the cubic crystal and I run a shear wave along a face diagonal, it is as if I am moving the ball bearings like this. There is essentially zero shear stiffness. Got this? The shear wave speed along the face diagonal is almost zero in many bcc materials. It is not known in plutonium but every other bcc material has maybe 5 to 10 times lower shear wave speed in that direction than it does along an edge.

That means that all the normal modes of that system are lower in frequency associated with that mode. That means it has lots of entropy, lots of states; if kt is large, your average quantum number is large. If the quantum number is large, the entropy is large.

This system has higher energy but it has this soft mode, which means it is very probable -- I mean, you can get tons of them, millions of configurations. The theory that I am working on now, which appears it is going to yield, is that this phase is strictly a result of the fact that we have a lot of entropy in that phase because it has soft elastic constants, so it is more probable and it forms.

There are 20 years of literature claiming that the F electrons in plutonium do something extremely weird to stabilize this phase and they have to do very strange things that are not probable and they have not done this computation. We are about to do this measurement completely using neutrons and resonant ultrasound and get the numbers.

To give you an idea of what the numbers are, TS -- over here -- the best we can compute from our preliminary dispersion curves is about 730 meV/Kelvin. The latent heat is 20 meV/Kelvin. That means that a change in the average sound speed enough to make the log of that sound speed change by 20 parts out of 730 could completely account for the phase transition of plutonium without invoking a single piece of weird F-electron physics and band structure, and I think we are going to pull that off.

[Transparency]

I am going to show you just one thing here and then we will go on to electronics. These are a bunch of measurements -- actually, there are more now -- that we have made on polycrystal plutonium samples using resonant ultrasound. The only measurements that ever agree with each other are ours on different samples from time and one other, Hassel Ledbetter at NIST, who made 7-mm-long single crystal at Rocky Flats before Rocky Flats turned into a Superfund site.

Those numbers and ours agree and everybody else is scattered all over the place, so we are getting a handle on it and I think we will be able to put this to rest using acoustics to do, I think, first-rate solid-state physics.

Okay, that is enough for that.

[Transparency]

I now have color, again. I am going to talk a little bit about instrumentation, noise, and stuff like that. A lot of this was supported by the National High Magnetic Field Lab in which I head an instrument design team. For some strange reason they are actually paying me to build electronics, something I enjoy a lot.

[Transparency]

The problem with noise interference, et cetera, is that you can only make things worse. That is the best you can do. I did not put the noise current in here. Anything you do will make it worse. Your job in getting data from your system into your hard disk is to make it only a little worse and you have to decide how. Tom has covered this stuff beautifully -- I am going to talk about that.

[Transparency]

Let's go back to Tom's picture of op-amps, for example. I am going to show you one way of screwing things up. It is amazing how similar our viewgraphs get to be. Here is a bipolar op-amp, CLC-425. This has 1.05 nV prHz, but it has a lot of current noise. Here is a JFET op-amp. It has 6 nV prHz input noise but almost no current noise. We are going to hook it up to a resonant ultrasound transducer in a second, but let's look at the curves first.

Here is the voltage noise, 1.05 nV prHz. It is flat with source resistance. Here is the current noise. It is just the noise current into the source resistor, so it proceeds linearly with source resistance.

Here is the Johnson noise from the source resistor, so where do you want to use this op-amp? It would be really nice to have the noise dominated by the Johnson noise, and that means, as Tom correctly points out, you do not pick it here, you operate anywhere from the point where the Johnson noise is bigger than the voltage noise to the point where the Johnson noise is bigger than the current noise through the source resistor, so I have a huge wide range for this very fine bipolar operational amplifier. It looks cool.

Now you go to a JFET input op-amp, an LT-1169, which is a particularly fine high-speed one. I do the same plot. Voltage noise is much bigger, 6 nV prHz. Current noise is really tiny, so I can off to 10 gazillion ohms and still be okay.

I now have a very long range where this amplifier is just great. Which one do I use? Six nV prHz or 1 nV prHz with the transducers I am using? You had better know what the transducers do, because here is what happens. I did not actually do this, I computed this, and I knew what to use beforehand, but I made this viewgraph by substituting one for the other, just to see what would happen.

[Transparency]

This is a 10-pF transducer into one of these op-amps. Here is what happens. White noise is good. Yes. Here is what the system produces with no signals. It is just wonderful and that is a random spike, somebody turned on a light switch, or something like that, we pick up interference, so we get white noise.

Now we hook up a resonator, a little sample, with that very low 1.05 nV prHz bipolar op-amp. Here are the resonances -- pretty crummy looking compared to the 6 nV prHz but no current noise JFET op-amp. So there you go. You can see this had lower voltage noise than this but the system noise overall, because I took care to compute impedances and current noise, I do lots better with that.

But if I am looking at 100-ohm source impedance I am going to win with this one. Anyway, just a graphic illustration of what to do.

MR. GLADDEN: What was the source impedance for that one?

DR. MIGLIORI: It is a capacitive-looking transducer, a lithium niobate single-crystal piezoelectric transducer, so if you look here from 20 to 140 kHz, the impedance is changing linearly with one over frequency.

Pulsed magnets: I need to get electrical resistivity out in 16 msec. Resonant ultrasound: I need to make, say, 600 lock-in amplifier measurements at different frequencies over the course of 10 seconds, the same measurement problem. The usual approach is to use a lock-in amplifier.

[Transparency]

What happens with a lock-in is that I have a signal with noise. I am going to walk you through it so you know just what it does. The first thing that is usually done is it goes through some preamplifier, which bandwidth-limits it and thereby does a little noise reduction, but that original initial preamp will have no effect on the final output noise of the lock-in.

The reason that it is used is simply to prevent the front end from overloading, because what is going to happen later is we are going to bandwidth-limit it down the line at a level much tighter than the preamp does, so the preamp is only limiting.

We have a signal now and what we do is we take that signal, we multiply it by an infinitely clean sine wave, point by point instantaneously multiplied, so that I get out a signal that now has frequency components at zero and twice the original signal frequencies -- that is the mixer or the synchronous switch, whatever you want to call it.

Then I RC-filter that and I get my output signal. Tom described it earlier, I shifted all my information up to the lock-in frequency by mixing it with itself, if you will. So I took information, say, at whatever the lock-in frequency was, a kilohertz, and I moved it down to DC, so I am measuring at a kilohertz. I do not have the 1/f noise problem and I do not have interference problems because I can narrow-band the system at a kilohertz, and then I shift it down to DC by mixing it with itself, and if I did the mixing right, I did not make too big a mess, and I am done.

I have a DC signal now that is proportional to the AC amplitude of the frequency that I want to measure and the bandwidth is simply an RC time constant now, because I am filtering around DC I do not have to make a fancy bandpass filter; all I need is an RC filter and the time constant tells me the filter bandwidth.

For example, a 1-second time constant has a 1-Hz bandwidth on the output. A 10th of a second time constant is a 10-Hz bandwidth, approximately. It is a way of making an AC measurement and easy-to-do AC filter, but there is almost no reason to use these things any more.

[Transparency]

Here is why. Who does not have a medium-sized digitizer in his lab? Everybody has got digitizers. Somehow you have to take data and put them into a Bill Gates program, except for Jay, who is smarter than the rest of us. You have to get data into your computer, so you are going to digitize something at some point.

The point is that you should digitize it right away. For example, if this is a sine wave and I make a clock generator that generates the sine wave at the frequency at which I wish to do the measurement and clocks the digitizer at an integer multiple of that frequency, then I get dots like this.

If I want to implement a lock-in function, I can software multiply these dots by a sine wave and then average them. Because this is synchronous, I am digitizing at a multiple of the fundamental frequency, it does not matter where I start and stop. After one cycle I have exactly the same number of data points of this value, this value, and this value, so when I average this way, I have a time constant, that is, the number of cycles I choose to average, that settles exactly at the end of my digitization process. I do not have to wait for an RC DK.

In addition, I am taking N digitizations per cycle and the digitizer resolves M bits, my result after I sum and average all of these is M plus N bits of real precision if I do it right, so I do not have need that high a number of bits on the digitizer. Signal to noise improves by the square root of N and I have every last piece of information in the signal when I have done this, which means, later, I can change the phase, the time constant, but I can do other things.

[Transparency]

By the way, here is a conventional lock-in settling with an RC filter and here is something where I have simply digitized. Actually, it is settled right here, but I have shown it settling fully after exactly one time constant in case I used an odd number of digitizations on it. That is the output of this digital scheme to do this.

[Transparency]

To make it more precise, here is a block diagram of the lock-in. Usually we have a low-level signal plus noise coming to a preamp, some limiting and filtering -- for example, the Stanford Research lock-in has an analog-to-digital converter -- it digitally multiplies by sine and cosine, it does a digital RC-type low-pass filter and software. Then it does a digital-to-analog conversion, and spits out an analog signal, which you then hook up to your digitizer and digitize, again. (laughter)

Here is what we are doing now and this is exactly how the resonant ultrasound systems work as well as the pulse-magnet ones. Remove this line, it does not belong here. Low-level signal plus noise goes to a preamplifier, which we digitize. We also have a sine-wave synthesizer that generates the sine wave and clocks the digitizer at an harmonic of that. Those go straight into the computer.

I have only one unavoidable analog-to-digital conversion in the system.

[Transparency]

Here is what we have actually done. You will see on this that we have digitized a signal synchronously. The second step: We read data into the computer.

[Transparency]

Listen to this. I have digitized the data with a digitizer of reasonable speed. Before I send it to the lock-in function in software I can clip it and limit it, I can get rid of spikes in a time domain that you cannot ever do in a lock-in. The first thing I do is time-domain filter it. Then I multiply it by the reference signal, which exists only in software at that point, and I begin the lock-in process of detection and then I filter it.

So I can now do both time domain- and frequency domain-filtering and if I do not like the RC time constant of the phase or the limiting, I can go back and do it, again, without having to reshoot the magnet or retake the data, if I choose.

So here is a pulse-magnet shot in a 60-T pulse magnet, 600 Gauss, in which we are trying to measure resistivity. The magnet comes up to 600 Gauss, then drops, and here is the magnetoresistance. Here is the raw signal with the sine wave in it. Here is the detected signal after we have multiplied it in software.

[Transparency]

Here it is after applying a Bessel filter to the output -- boom, like this -- so these are now resistivity data in a pulse magnet.

[Transparency]

This is a little bit more detail on what is going on. Here is the raw sine wave. We have just multiplied it by the artificially generated sine wave and here it is after we have put on a Bessel filter. We can just play games after the shot as much as you want. You see here all I have is a preamp and a digitizer. I have thrown away the lock-in amplifier, and I have done the rest in LabView.

[Transparency]

There is another interesting aspect to it. When I RC-filter, there is a time shift. Your RC filter acts by causality after the data, but if I take all the data and go back and filter them, I do not have a causality problem and I can actually take -- for example, the red is an RC- filter function and it is delayed in milliseconds over the real signal, but I can center my filter at zero time instead of having a delay and remove that delay, so the blue shows the superconducting transition of BiSrLaCuO, a high-Tc superconductor, in a big fat pulsed field.

So now I have the fields correct, because I do not have this time delay associated with the RC filter that does everything.

[Transparency]

Just for fun, this is not a Stanford Research but this is another popular lock-in. Here we took data using just the digitizer and the software functions. Preamps were identical in this one. This is the output of our system in which we took all the data and then manipulated them afterward.

This is the output of a digital lock-in, a commercial one. What happens is it tends to semi-autozero itself at random times, so you get these level shifts -- they are small -- and then, when it spits out its analog output at the end of all this processing, there is digitization noise that is not properly filtered, because they are trying to have a very fast time-constant output, so you get this very high-frequency stuff.

In fact, the black and the red traces have exactly the same time response for real signals, but the commercial one has a lot of excess hash in it

[Transparency]

This is resistivity and the Hall effect on a Ag_2Se sample in a pulse-magnetic field. As far as we can tell, these data, taken in 16 msec in a magnet that has been pulsed by a capacitor bank about half the size of this room -- we closed an ignitron switch, we dumped 20,000 Amps at 10 kV into this magnet this big surrounding the experiment and we are trying to measure microvolts off the sample.

Here it is. These data are as clean as if you took them in a superconducting persistent magnet in the lab with shielding around it, because we can attack almost all of the problems.

[Transparency]

Just for fun, this is the dual digital synthesizer that right now is only a serial port interface that generates the two clock signals that we then feed to any commercial digitizer that we like, so we have sine- and square wave-outputs at two frequencies that are clocked and phase-locked together and that is what we are using right now for most of this stuff.

I think I am going to quit there. Thanks.

MR. GLADDEN: Your digital lock-in, the software is programmed in LabView?

DR. MIGLIORI: Yes, just because my technicians are really familiar with it.

MR. GLADDEN: But it is a combination, you have some hardware in there?

DR. MIGLIORI: That is right, we are driving the synthesizer with LabView, we are driving the digitizers with LabView and, of course, I have to do an in-phase and quadrature, multiplying afterward, and I have to keep track of the starting times to keep the phase correct after each shot, but all of that is pretty straightforward.

The synthesizers are available from the guy I gave the circuit diagrams to make the boards for Los Alamos -- it is in Albuquerque -- I think they are about $1000 apiece. I think if anybody wanted the LabView drivers, we would give them away to anybody who wanted to use them. We do not care.

DR. GARRETT: The National High Magnetic Field Laboratory, I thought that was in Florida.

DR. MIGLIORI: There are three campuses: Los Alamos for the pulsed magnets, Florida State for the D.C. magnets, and the University of Florida for the NMR magnets.

That experiment in the pulse magnet uses somewhat less than a penny's worth of electricity when we pulse it. To do the same measurement in the FSU bitter magnet, which is the DC one, they are running 20 mW for about an hour.

It is really fun to be able to do some experiments where the pulse-magnet data are perfect. There is no reason to have to run these huge power supplies unless you are doing a thermal measurement. Next year I will show you how to make thermoconductivity measurements to 1% in 10 msec on anything. I did not have time this time for that, but these are the three omega techniques, which we are also working on.

Often there are experiments that cannot be done in a pulse magnet, but the number is decreasing. We still cannot figure out how to do resonant ultrasound in a pulse magnet.

DR. GARRETT: When you have this godawful magnetic pulse, how come it does not just induce resonance and burn up your signal processor before you get to do the signal processing?

DR. MIGLIORI: This is a fair amount of art in attempting to keep the open area, loop area, very small on the leads that come in. The samples are very small. You run things parallel to the field, the wires are parallel to the field coming out. There is a lot of black art to it.

In the end, we do get 1-V slowly varying and clean on top of 100-μV transport signals at 150 kHz. We are running a lot of these at fairly high frequencies.

Thanks.

ATMOSPHERIC AND METEOROLOGICAL ACOUSTICS

Kenneth E. Gilbert
University of Mississippi

DR. GILBERT: I started out going to graduate school right before the Vietnam War and that was a golden era, but during that period it got to be not so golden. I had a friend from the Applied Research Lab at the University of Texas who showed a viewgraph that I want to show. It might have been in Physics Today or it might have been in Science magazine. I have never been able to find it since, and I have drawn a crude representation of the cartoon.

[Transparency 1]

If you can send me a nice photocopy, I will reward you handsomely. Here is how it went. You have these two guys working on a car. It is a young guy and an older guy. One says, "Gee, pop, how did you end up as an auto mechanic?" He says, "Well, son, I started out in quantum mechanics but I made a wrong turn somewhere."

I started out in quantum mechanics but I made a right turn and ended up in acoustics. I started out working for the Navy and I had some fun but, really, the most fun I have had has been over the past decade working in atmospheric sound propagation. The reason is there are just a lot of data out there. Ocean experiments are very expensive and there are not a whole lot of data. In atmospheric acoustics there were plenty of data going back to the 1960s and no one had an explanation for it.

During the daytime, the ability to predict sound levels was limited to about 100 meters or so and there was no explanation for data that had been measured at that time 30 years prior.

[Transparency 2]

The story I want to tell today is progress in the past decade in understanding daytime sound levels. I originally started to tell about everything important in outdoor sound propagation but I realized that not only was that dull but I could not do it in three hours.

[Transparency 3]

What I am going to do, basically, is tell a research story. The last bullet, “A Research Story,” is what I am really trying to get to but I have to make clear what we are trying to find out and along the way I will have to tell you about the atmospheric or turbulent boundary layer. My time at Penn State was very valuable to me, because I became a close collaborator and colleague of John Wyngaard, who is a world-class meteorologist, and a little of what he knows rubbed off on me.

To understand the research story you will also need some information about sound propagation. An important tool that has come into widespread use in this past decade is the parabolic equation and I will try to give you the flavor of that. Then I will tell the research story.

[Transparency 4]

Let me pose the first question. The question is what are the mechanisms that control sound levels in the daytime acoustic shadow region? We are going to proceed with the most useless equation in acoustics, the wave equation, but for me and for anyone interested in propagation it is the most important equation.

What we are really doing is applying acoustics to propagation in the real atmosphere. The theme of this talk, is that nature has given us an atmosphere, it has been there a long time, and it has certain systematic properties.

One of the things I will want to emphasize is how tightly what you observe in acoustics is controlled by what is in the atmosphere. If you try to invent your own atmosphere you will eventually go astray. We tried to stick to that theme but we got a little bit off track, but some colleagues from the Ecole Centrale de Lyon set us straight and once we got back to what the real atmosphere is, things made a lot more sense. That is part of the story I want to tell.

[Transparency 5]

I will be talking about daytime, primarily, but to give some perspective I will talk a little bit about nighttime propagation. When you have daytime propagation you effectively have a sound speed that decreases with height, so that if you imagine a wavefront coming along like this, the bottom half is going faster than the upper half and it turns upward, you get upward refraction.

At night you have the sound speed lower near the ground and so part of the wave is refracted downward and bounces, so you get a duct here. This process of ducting and producing quite long ranges is a fairly well-understood thing.

The levels here in the shadow zone were not understood. Various things were tried, such as surface waves and some phenomenological approaches. The fact that it was maybe due to turbulence had been talked about by Gilles Daigle, but no one had ever really tried to make realistic calculations that you could compare with data.

Mike White, a former graduate student at Ole Miss, had done his Ph.D. thesis and he had put together the first parabolic equation that was actually used in atmospheric acoustics. But Mike left, took a job, before we got to do the turbulence part, so my colleagues, Richard Raspet, Xiao Di, and I picked up at that point and wanted to ask the question: If we take realistic turbulence, can we explain what is going on in the shadow zone?

[Transparency 6]

This effective sound speed that was on that previous picture, to a good approximation, is composed of two parts. We have what is called the adiabatic sound speed. This is a very well-known expression and you can write it in a simpler form, where you have a reference sound speed C_0 and a reference temperature T_0, but the main thing to note is that it varies as the square root of absolute temperature.

For example, if you pick C_0 as 331 and T_0 as 273, that is 0°C, then this expression works just fine. The effective sound speed has, in addition, the component of wind along the direction of propagation. The quantity $\hat{n}_p$ is a unit vector in the direction of propagation, so the effective sound speed is the adiabatic sound speed, the still-air sound speed, plus the advection due to wind.

When you have a situation where you have an upward-refracting profile and wind, if you go against the wind, the wind will refract upward, so you get even more upward refraction. If you go cross-wind nothing happens very much; it then is just the adiabatic sound speed. When you go downwind there is a battle between upward refraction and downward refraction, and sometimes the wind will win, sometimes the temperature will win.

You have four directions, in general. You have upwind and cross-wind (twice) and downwind, so in three of those four directions you are going to have upward refraction. That is basically what is going on in daytime.

[Transparency 7]

Dave Swanson, at Penn State, did a nice measurement that illustrates this phenomenon, diurnal (day by day) variation. He measured the temperature a 10th of a meter and a half-meter off the ground. During the daytime the temperature goes up and it cools down at night. It goes up in the day, cools down at night.

If you look at the associated sound-pressure levels, they are high at night and down in the daytime by about 10 dB, so as the temperature goes up, you get upward refraction, and the levels in the shadow go down. It is these levels that we would like to predict from a really physical basis. This is only a 10-dB difference. This is 54 Hz at 450 m. If you go to higher frequencies and longer ranges this can easily be a 20- or 30-dB difference between day and night.

So that is the phenomenological view of what goes on.

[Transparency 8]

In order to begin to put some physical basis for all of that, we need to talk about the atmospheric boundary layer. Sometimes that is called the turbulent boundary layer and it is basically the first kilometer of the atmosphere.

[Transparency 9]

Let's look at a cartoon to get our perspective here. On your notes I have "unstable" up here and Michelle Swearingen said that was confusing, so I moved it down here. This is a measure of the edge of the turbulent boundary layer. Above that we have free space, which is called the troposphere. It goes up to 11 km. Then you have the stratosphere, where the high winds are, and then you have the mesosphere, and then the thermosphere, so it goes up to 150 to 200 km, so in this lecture we are looking at the bottom 1% or less of the atmosphere.

But this is where we live and this is where sound happens unless you go to infrasound. In the region of a few hundred hertz, this is where sound happens. The atmospheric boundary layer is pretty much driven by the sun. When the sun comes up in the morning, you start heating the ground. It really does not heat the air very much, it heats the ground, and then the ground heats the air and it becomes unstable and thermals start rising. If there is a wind blowing, as there usually is, you start getting these big atmospheric boundary layer eddies. This is the source region.

Then it couples through nonlinearity to smaller and smaller eddies until it gets down to a millimeter level. This energy that is being put in at hundreds of kmeters disappears at a level you cannot even see. The viscosity at a millimeter is sufficient to soak up the energy of the eddies. There is also energy loss of the mean wind due to friction with the ground.

Clouds generally are right above the top of the atmospheric boundary level here, so when you see these flat-bottomed clouds, you know they are sitting right on top of the turbulent boundary layer. If you ever fly in the daytime, it is very obvious when you pass from the turbulent boundary layer into the free atmosphere; things just suddenly get smooth.

What happens when the sun goes down is you start getting radiation from the surface into space and the ground starts cooling off and it no longer has a heat flux to drive it and it collapses pretty rapidly and you build up a stable layer near the ground. This stable layer is what drives nighttime propagation. This turbulent boundary layer is what drives daytime propagation.

[Transparency 10]

There is a particular phenomenon that is called the adiabatic lapse rate. I asked a meteorologist, if I took an insulated tube of air a kilometer long and stood it on end and came back a year later, what would be the temperature distribution inside of it? He said the adiabatic lapse rate, but that is wrong. It would be a constant temperature. It would have a pressure gradient, but it would be a constant temperature.

So why does the atmosphere like a gradient of 9.8° C/km? Does anyone have any idea why it is you can have this?-- It is a neutral atmosphere but it is hotter near the ground than it is up here -- why is that? How can that be neutral?

DR. MIGLIORI: It has to do with the circulation or something.

DR. GILBERT: That is the key to the adiabatic lapse rate. What happens is if you have a parcel of air that goes up, it will cool adiabatically. If it goes down, it will heat adiabatically. What you get is about 9.8×10^{-3} C/m. So this -- right here, the adiabatic lapse rate, -- provides a background for which a heating or cooling parcel of air stays neutral. The atmosphere always has mixing processes essential for setting this up. So if you have some turbulent mixing, if you leave it alone, it will go to this neutral condition, which is 9.8°C/km.

The reason you do not really get an adiabatic lapse all the way to the ground is that the mixing cannot keep up with the heating and cooling of the ground. When the sun comes up, you start getting heating near the ground and it gets to be a bigger gradient than you have with the adiabatic lapse rate and the air near the ground, then, really does become unstable.

If the sun just stayed in one place, eventually this would slide over to here so there was an adiabatic lapse rate all the way down. The temperature gradient gets steeper and steeper through the day, but when the sun goes down, it starts cooling off and, again, it cannot keep up and you get this part in here that is going backward, but this deepens over time. If the ground stayed the same temperature forever, then this thing would eventually slide back over to here and you would have the adiabatic lapse rate all the way to the ground.

That is a particularly important thing and it pretty much is what drives the dynamics of the atmosphere.

[Transparency 11]

In particular, there are always fluctuations and that is partly what we will talk about, but if you look at the mean adiabatic sound speed, it does the same thing that you were seeing with the temperature since it varies as the square root of the absolute temperature.

As the sun comes up, you start getting this unstable air, it is more than adiabatic, and you have upward refraction. In the afternoon you get the biggest heating and you get very strong upward refraction. As the sun goes down, you get cooling and that refraction gets weaker. Then, as it cools at night, the upward refraction turns into downward refraction and you can get long-range sound propagation.

From a practical point of view, the nighttime levels are a lot more important in noise control than in daytime, because it is at night where you have noise problems, but I am not going to talk about that. I am going to talk about daytime, because that is what we have been studying for the past 10 years.

[Transparency 12]

What about the wind? This is a figure from one of John Wyngaard's recent papers, a little tutorial, that shows what the wind looks like. It also shows what is called the potential temperature. Since potential temperature is a very common term in meteorology papers, I will tell you what it is.

This profile is pretty much logarithmic. It has to go to zero near the ground. This is the temperature in the day, and this is the temperature at night. This peak in wind speed is called a nocturnal jet. In general, we can think of the wind profile as being logarithmic.

If you notice, we have a theta here that does not change much, except at the very bottom. Theta is the potential temperature and here it is the mean potential temperature. The potential temperature is just the real temperature with the adiabatic lapse rate taken out, so the potential temperature is a constant when you have a neutral atmosphere and, when it gets over here to the left, it is stable.

When it gets over here to the right, it is unstable.

These are little thermals. These are big eddies being created. Actually,air is entrained from the free atmosphere and some of the turbulent atmosphere is injected into the free atmosphere. The height of the nocturnal boundary layer is a couple hundred meters.

The height of the daytime boundary layer is maybe as much as a kilometer.

That is John Wyngaard's picture and that is what potential temperature is.

[Transparency 13]

What I would like to do now is show you what this beast really looks like. This is a picture of aerosol concentration measured with a laser device called Lidar. It is using the aerosol concentration as a marker of what the atmospheric boundary layer looks like. It is up to about a little bit over .6 km. What you can see is all those pretty smooth lines are really quite irregular and ragged, so this is really kind of a mess.

You can imagine that there is nothing really smooth in there, that there are lots of fluctuations inside the turbulent boundary layer. Surprisingly, the computational methods that meteorologists use today can pretty much capture that.

[Transparency 14]

This is a large eddy simulation calculation done by Martin Ott at Penn State. We were doing electromagnetic propagation. If I have time, I will say a little bit about what we learned. This is the moisture content as a marker of the turbulent boundary layer.

You can see it shows pretty much the irregular nature of the volume and the edge of the boundary layer. The resolution here is 20 m. Actually, there is structure down to a millimeter but computers are not up to that yet. We can sort of fake it out but we cannot really calculate it, not for the Reynolds numbers that are about 10,000.

Anyway, I am beginning to give you the personality of the turbulent boundary layer. Jim Chambers and some of the young guys at NCPA went out to the Oxford Airport with some sonic anemometers. There are three components, they can measure the x, y, and z components of the wind and they can measure the temperature and they sample it 10 times a second, so you can really follow in time the kinds of structure you are seeing in these snapshots.

[Transparency 15]

This is what we measured about a meter off the ground. You are beginning to see now the strange behavior of the wind. This is the horizontal wind.

This is the vertical wind. It has a mean of about zero, and why do you think the vertical wind component has to have a mean of about zero?

PARTICIPANT: You would run out of air.

DR. GILBERT: Right. You do not get flow into the ground, so it has to average out to zero.

The horizontal wind is really fractal. You have large scales with intermediate scales, with fine scales, so when you start talking about means, you really have to say over what time period. Obviously, if I make an average of the mean wind over a period of a year, I might have a hurricane in there somewhere.

Generally, the longer you average, the bigger the average and the bigger the variance. We are going to be interested in scales of about 10 or 20 m and less. What we are really interested in is averages over about this period of time (tems of seconds) -- oh, I should say, excuse me, if you look up here, you see an average horizontal wind speed, more or less, of about 2 m/sec. So in 10 seconds, 20 m of air goes by you.

We are using something called the Taylor frozen hypothesis. If you imagine a coat hanger wire with lots of wiggles in it and you just had a little slit that you could look through and you pull the coat hanger wire by, you see this thing going up and down. It is just a mapping in the time of the structure in space. That is Taylor's frozen hypothesis, frozen turbulence hypothesis.

It is really not true but it is good enough for the statistics. If you try to take it literally, you will find that it does not really work so well. Nevertheless, what we do in order to take this time structure, we multiply it by the mean velocity and seconds become meters. If I multiplied all these numbers by two, I could put meters down here and this would give some kind of representation of the spatial structure in time.

What we want to do is look at the wavenumber components in that spatial structure. If you Fourier transform the wind data directly, you will get the frequency components. We want to discuss spatial components, so we assume the turbulence is frozen and convert the time series to a spatial series.

[Transparency 16]

The temperature looks similar. If we are interested in things on the order of 20 m and down, we are talking about maybe 10-second averages, so we want to look at the mean and variance over periods of about 10 seconds.

It turns out that, numerically, the standard deviation in wind and the standard deviation in temperature is about the same number, about a half-meter per second and about half a degree. The means can be whatever the means are.

[Transparency 17]

What we want to do now is start to get some estimates of how big the mean is and how big the variance is. The daytime mean, you could measure it directly, but you really do not have to. There is a similarity theory by Monin and Obukov. If you measure two temperature values near the ground, you can apply the similarity theory and get a pretty good representation of the mean temperature.

As I have said already, the velocity is logarithmic with height. This little z_0 is a small number that represents the thickness of the dead air, so in the dead air -- it is not really dead, it is just not turbulent -- you have some V_0. At z_0 this would be 0. Anyway, we have a logarithmic variation.

What we want to do is represent the temperature in terms of a mean and a fluctuation and, similarly, the component of wind is a mean and a fluctuation. I want to give you some orders of magnitude, so you can have something to think about.

The variation of temperature with height is going to be 5° or 10° Kelvin or C. The variation in height of the wind will be 5 or 10 m/sec. The RMS fluctuation is about 0.4 or 0.5 Kelvin for the temperature and about .5 m/sec for the wind. If you take the expressions I showed you for the adiabatic sound speed and you expand them linearly, you will get your reference sound speed plus 0.6 times the temperature difference, and this is with height.

When you take 5 or 10 and multiply it by .6, you get 3 to 6. Up here, with the wind, we have just the same thing, 5 or 10 m/sec. You can see these are comparable. The temperature and the wind contribute about the same when you look at the mean values, but when you look at the variance or the standard deviation, this 0.6 in front of the temperature does big things.

If you put in 0.4 and square it to get the variance and multiply it by 0.6, you get 0.058, where this one 0.5^2 is 0.25. The wind contribution to variance is about four times that of the temperature, so when you talk about the variance in the effective sound speed, it is dominated by the wind fluctuations, whereas the means have about equal contribution from wind and temperature.

This is an important fact, because it really says that all you really need is the fluctuation in the horizontal wind component, because the fluctuations in the effective sound speed are dominated by the wind.

[Transparency 18]

What we want to do is just make some simple linear expansion of things and take this temperature and wind information and convert it. I will not go through all the mathematics. All I am doing is making linear expansions, saying that delta T is small compared to the mean value.

If you expand this out, you will see, not surprisingly, that the mean effective sound speed is the mean adiabatic sound speed plus the mean horizontal sound speed. Then you get, over here, the fluctuation in sound speed from temperature and wind fluctuations. The whole coefficient of delta T is about 0.6, as I showed before. The last term is just the fluctuation in the wind itself. So we have a mean component and a fluctuating component for the effective sound speed. That sort of view will carry through the rest of the talk.

[Transparency 19]

Finally, what we want to do is not use sound speed but use index of refraction, so we need a little bit more linear expansion. This is the definition of the effective wavenumber. I can just put in a reference sound speed and this, then, is the effective index of refraction, $n_e = c_0/c_e$..

I have the effective sound speed, which is composed of a mean and a fluctuating component, and all I want to do is linearly expand the thing, so we get this expression for the mean index of refraction, and we get this expression for the fluctuating part.

If you go back and take the expressions for the mean effective sound speed and the fluctuation in it and plug into that, you will find that the mean index of refraction is just your reference sound speed over the means for the adiabatic sound speed and the wind speed.

This fluctuation in the index of refraction is this quantity times the wind fluctuation, and this quantity times the temperature fluctuation. To give you some feeling for what these things are, this number, the mean index of refraction, is not much different from one.

This number right here, the fluctuation in the index of refraction, is going to be on the order of one part in a thousand. The fluctuation in the index of refraction is not a very big thing but it has big effects when you put it into propagation.

The important thing at this point is to get a feeling for what these fluctuating components look like. Remember, if they fluctuate in time, it means they have some complicated structure in space that gets advected by you. If you have noise, what is one of the primary characteristics that you want to know about the noise? You have this wiggly bunch of stuff like Tom showed. What is one of the primary thing you want to have in describing the characteristics of that noise?

You might think in terms of the autocorrelation function, but what else is more commonly used? The spectrum. The spectrum of noise tells you a lot about the nature of the noise, 1/f noise and other types of noise.

What we want to do with these fluctuations is to talk about the wavenumber spectrum for the fluctuations. The wavenumber spectrum tells what the fluctuations look like in space.

[Transparency 20]

It turns out that there is a part of that spectrum in here called the inertial sub-range, and it can slide up and down, it can go up and down in strength, but the slope is universal. The slope was predicted by Kolmogorov in 1941 through some scaling arguments. No one had ever really measured it at that time. He did not really calculate it, he just was a bright man.

This slope is a -11/3 in three dimensions, it is a -8/3 in two dimensions, and a -5/3 in one dimension. In fact, the measurements we make are always one-dimensional. We stick a sensor up, the atmosphere blows by, and you get all these wiggles, so you have to make some assumptions, like horizontal isotropy or three-dimensional isotropy to go to bigger dimensions.

The spectrum in this part is called the Kolmogorov inertial sub-range spectrum. It is really not a spectrum, it is just a piece of the spectrum. Up here, where this thing rolls over, in this region in here is those big huge eddies that go up a kilometer. That is the source region.

This region down here is where you have eddies the size of millimeters and viscosity soaks up the energy. You are putting in energy here in these huge big eddies, but they couple, because it is a nonlinear process, to smaller and smaller eddies until they are down to the size of a gnat.

What Kolmogorov realized was that in order for the cascade of energy -- and this happens only on average, if you average out the spectra over a period of time, this is what you get -- what Kolmogorov realized was that if I am putting in energy, then for the cascade of that energy through the various scales to be uniform, there has to be a special spectrum for that to happen.

He deduced that it would be this. This is what is required for this energy put in, in the large source scales. In order for it to cascade and not build up anywhere, but cascade smoothly down to the place where it is dissipated, this is what has to happen. In a sense, it is kinematic.

This break point on the left is called by the Russians the "outer scale," and this break point is called by the Russians the "inner scale." In the U.S., in particular, this is called the "integral scale," because you can define it in terms of an integral over the autocorrelation function, and this is called the "Kolmogorov microscale." The inner scale is millimeters. The outer scale is hundreds of meters. That is the generic view of the turbulence spectrum and that is what nature gave us. Unfortunately, when we started our work, that is not what we used.

[Transparency 21]

This is what I was talking about. This is just the Kolmogorov line, -5/3, -8/3, -11/3, depending on the dimension. The actual spectrum rolls over at this end and at this end. We used, for reasons I will explain, a Gaussian spectrum. That is what seemed to be observed by acousticians.

Meteorologists observed this (actual spectrum) and at that time this (Gaussian spectrum) is what acousticians took as gospel. It happened to fit, in a small region, the actual spectrum. The actual spectrum is what nature gives us and that is the theme, in my mind, of atmospheric acoustics, to look at the atmosphere that nature gave us and deal with it and not invent our own atmosphere.

In underwater acoustics, people for many years invented their own ocean, because it made airy functions, special functions, nice and made the integrals nice, but they were always led astray, always. I do not know of any instances where eventually they were not led astray, and it is the same in atmospheric acoustics; you get led astray when you deal with something that is not what nature gives you.

[Transparency 22]

Now I am going to just define some terms and then I will finish up this section.

[Transparency 23]

I want to talk about some of the basics of propagation. What we deal with is the wave equation and we use the effective sound speed. In the wave equation are some things like geometrical spreading, refraction, diffraction and scattering, and attenuation and absorption.

I will not talk about this (attenuation), because for the frequencies and distances we are concerned with attenuation is not a huge factor. At higher frequencies it can be a huge factor, but we will talk about just the first three topics to show you what I mean by those terms.

[Transparency 24]

Geometrical spreading is basically a conservation-of-energy concept. If we have a point source and you look at the energy flowing through a spherical shell, this is in free space, in order for the same amount of energy to flow through a surface that is increasing as r^2, the intensity has to go down as r^2 and the pressure will go as the square root of that, so the pressure is falling off as 1/r and the intensity is falling off as $1/r^2$ in spherical spreading.

In the nighttime situation, where the energy is trapped near the ground, I drew a cylinder here to suggest a cylindrical surface that these rays are passing through. This surface is increasing with radius as r, so intensity goes down as 1/r and pressure goes down as 1/square root of r.

In the general case where there is bending, it is not so easy to get the pressure or intensity, because we have no general rule that works for arbitrary propagation.

[Transparency 25]

Another word that is important that I have been using already is refraction. As I said, it is easy to understand if you think in terms of the sound speed. It is higher down here near the ground. The bottom half is going faster, so it turns up. Here it is the reverse. The top half is going faster, so it turns down.

As long as this bending is spatially slow compared to a wavelength, you just get these wavefronts that go up or they go down, but that is not the whole story.

[Transparency 26]

When you have structure that is large compared to a wavelength, such as the mean profile, there will be a general refraction. When you have smaller structure -- this is supposed to be structure all out in here and you are thinking of turbulence as a huge mass of eddies or turbules (meteorologists call them eddies, acousticians call them turbules) – then something else happens, you get diffraction.

The point of it is that these things represent wavefronts. When you come into contact with stuff like an eddy, new wavefronts get created. Before, to describe what happens to a wavefront, all you needed was Snell's law. Everybody knows Snell's law. That describes the normals to a wavefront. With structures on the scale of an acoustic wavelength or smaller, you have to account for diffraction because, in reality, the wavefronts get split into more wavefronts. To describe that splitting in terms of physics, you have to deal with the wave equation. The wave equation accounts for diffraction as well as refraction and geometrical spreading.

I will talk about solving the wave equation in the next session, and I guess this will be a good breaking point. Does anyone want to ask any questions at this point?

MS. SWEARINGEN: You had mentioned that for the adiabatic lapse rate if the sun did not come out or it did not go down, the temperature profile would eventually collapse into that. Has anybody done a study on any part of the earth where you do have sunshine for days and days or nighttime for a long time?

DR. GILBERT: It is not so much what the sun does, but if the ground stayed at a steady temperature and it started out unstable, the mixing processes would go on until it reached a neutral condition, which is 9.8°/km. At that point, if you move a parcel of air up, nothing happens, it is neutral. It is going to cool off as it goes up and heat up as it goes down, but all the other air around it is at whatever temperature it finds itself at. That is the neutral condition.

The dynamics of the mixing are just going to work until that neutral condition is established. But that never happens. The ground is always heating up and cooling off, so this adiabatic lapse rate is always trying to catch up with what is going on, on the ground.

At night there is not nearly as much mixing, so that process of catching up is very slow and you get this big cool layer of stable air near the ground at night, and that pretty much defines the dynamics. The mixing at night is just mechanical turbulence. Mixing in the daytime is convective. When the air is unstable, the wind can just take it up forever.

DR. DENARDO: If the turbulence is associated with eddy motions, how can you have turbulence in one dimension?

DR. GILBERT: Tell me what you mean by that question.

DR. DENARDO: I thought you had the turbulence spectrum for three different dimensions. Maybe I did not understand.

DR. GILBERT: Imagine you have three-dimensional turbulence coming by and it is frozen in space and it just advects by you. The one-dimensional part says if I take a cut along a line I will get a certain spectrum.

DR. CRUM: I saw something very interesting yesterday and I presumed it had something to do with water vapor in the atmosphere. You did not mention much about water vapor. I presume the frequency here is an important thing, but when we were landing yesterday, when the flaps were down, there was a vapor trail coming off the edge of the flap. What happened was that the vortex coming off the end of that flap was a corkscrew, so I presumed there was some nucleation and evaporation and condensation there giving that vapor trail.

What is the role of vapor in all of this and can you explain why these things develop?

DR. GILBERT: Water vapor is not a huge factor in the acoustic sound speed. For electromagnetic waves it is everything. Water has a dipole moment at megahertz to gigahertz frequencies. Because it has a permanent dipole moment, it interacts with the electric field very strongly and controls the phase speed of the electromagnetic wave. Those calculations, the large eddy simulation I showed you, were for water vapor and it is nothing but a solution of the Navier-Stokes equations that has had the small scale filtered out so they can deal with it numerically. We did those calculations because we wanted to predict the index of refraction for electromagnetic waves, and that is almost totally driven by the water vapor content.

In acoustics, the water vapor content is not a big deal; it is a small effect compared to everything else at these frequencies, which is a few hundred hertz up to maybe a kilohertz. It is not a huge effect.

DR. SABATIER: When you have....six months. Is it just a long night in terms of sound propagation? It is just early morning propagation?

DR. GILBERT: That is a good question. I do not know if the ground would ever reach an equilibrium. It is always radiating into space and that is what happens at night, this energy is just going off into space as the ground gets cooler and cooler. I do not know where that stops.

If you could get the ground to stay at a steady temperature, if you had a kilometer-wide hot plate and held it at a steady temperature and could exclude the rest of the atmosphere, the mixing processes would go on until you got 9.8°/km, because then it becomes neutral. That is the boundary between stable and unstable, is this 9.8°/km. It is interesting because it looks unstable. You have hotter air near the ground. Why didn't it rise up? The reason it does not rise up is because the air it gets next to is the same as it is, so it does not have any buoyancy, positive or negative.

MR. GLADDEN: You were talking about the fractal nature of wind speed, so you have to be specific about the time slices you are looking at. What guides you on choosing what time slice to look at?

DR. GILBERT: What happens is, if you create the spectrum, it all takes care of itself. The way acoustics works is it selects out a piece of the spectrum. You have the spectrum versus wavenumber and if you take a slice, what you are interested in is the variance of that wavenumber interval slice, so your acoustics tells you what piece of the wavenumber you are looking at.

If you are trying to express a mean and a variance to somebody, if you are talking about this little slice, you are talking about a time period that you are going to look at because it is a certain range and scale. Obviously, if I average temperature over a full day, it is going to be different than if I average it over an hour. That is just the way chaos and fractals work; everybody is pretty much comfortable with that nowadays.

I will ring this bell in about two minutes.

DR. GILBERT: A question was brought up at the end of the session about how a parcel of air cools when it goes up by expanding and is compressed when it goes down. What I forgot to say is that there is a pressure gradient in the atmosphere and when a parcel of air moves down it is going to higher pressure and it gets compressed and heated adiabatically. When it goes to lower pressure it expands and cools adiabatically, so that contraction and expansion is driven by the pressure gradient in the atmosphere. I forgot to say that.

[Transparency 27]

I hope you have lots of coffee in you, because I am going to show some equations. I will try to keep this as cartoonish as possible. If I can explain things with cartoons instead of symbols, I will.

We talked about the basics of propagation, refraction and diffraction, and all that good stuff. It turns out that in about the early to mid-1970s a new approach was brought into underwater acoustics by Fred Tappert, and that was the parabolic equation approximation.

It was actually invented in the 1940s in an analytic sense by Leontovich and Fock, a couple of Russian guys who analytically wrote this equation out, and Tappert realized that computers of those days in the 1970s could solve this thing and you could do wonderful things with it.

What I want to do is give you an introduction to the parabolic equation approximation. I am not going to do it the way Tappert did, because today we have a better understanding of what is really going on than was apparent in those days.

I have Tappert's method written down here and maybe, if there is time at the end, I will go over it.

[Transparency 28]

First of all, I want to give you the flavor for what the parabolic equation is all about and why it is important. It has been in the atmospheric community, really, for only about 10 years. The question is what is so special about the parabolic equation in outdoor sound propagation?

The answer is that right now it is probably the only practical method for dealing with "range-dependent" environments. What does that mean? If you look at a sound speed, or any parameter that varies only with height, that means it is not varying horizontally, so you can think of this as a bunch of layers of plywood that extend forever this way.

When I was explaining this to John Wyngaard, we explained and explained, and he said, finally, "Oh, you mean it's like plywood." I said yes. We actually used that term in an electromagnetics paper, we got it past the referee, and it is now known in the electromagnetics community as the plywood approximation, so I guess we will start it in the acoustics community, too. When you have things that are just strictly layered, it is like a bunch of sheets of plywood.

That type of thing allows you to do conventional separation of variables and you reduce it to ordinary differential equations and you can use standard methods. Great. But, unfortunately, nature does not pay attention to the plywood approximation. Nature is very complicated in every direction.

You have a situation where, if you indicate a bunch of profiles that are changing with range due to turbulence, you have something like this, or you can have terrain that is not flat. For these types of environments that are range-dependent, we have to have a new approach. That new approach I will be calling the parabolic equation, but it is also known as one-way wave equations. One-way wave equations and wave extrapolation came out of the seismic community. We do not usually say wave extrapolation in acoustics but the seismic community still does.

To avoid all that, I will mostly be talking about the parabolic equation.

[Transparency 29]

What are the important points that we need to think about? First of all, conventional separation of variables does not apply globally in complicated media, but you can slice up the medium into little thin slices and apply it in there. If you want really fine slices, you are really working your head off to do it that way.

Another way would be to just take the equation and grid the world and solve it directly. If you sit down and count the number of points you need to do that, you lose your enthusiasm pretty rapidly. In two dimensions you will need something like 10 million to 100 million points. In three dimensions you are going to need something like 10 billion or 100 billion points. That is really kind of daunting.

If you use the parabolic method, what you are really doing is you are taking a starting field -- this is supposed to represent a Gaussian -- and you march it out and you are having to keep account of only these points, because the parabolic equation assumes one-way propagation. There are no echoes going backward, so it does not know what is out here until it gets there. You are just tracking maybe a couple thousand points vertically, and that is all you have to keep track of.

In 2D it is typically a couple thousand points. In 3D it is 100 thousand to a million points, but even a million points per range step on a fast P.C. is not too bad. People in the meteorology community routinely take time steps with 10^6 points per step. They now do their calculations on P.C.'s and they routinely will be advancing in time a million points.

I will show you at the end some 3D calculations. One of the main things they did for us was to show that the mean levels computed with 2D calculations are almost identical to the mean level you calculate with 3D calculations.

The way of thinking in terms of a marching solution is really the new idea. I say wave equation but I am really talking about the Helmholtz equation; we are really talking about harmonic solutions.

[Transparency 30]

If you take this wave equation here at the top and you assume that we have some arbitrary space-dependent amplitude multiplied by our harmonic term that has a space-dependent phase, you can write this in complex form and take the real part and then this amplitude times this phase becomes a complex amplitude and that whole thing I call P hat to indicate it is complex, so we want to take the real part of this.

Taking the real part is a linear operation and we have a linear equation, so we can just plug in that solution and go with it, and we find that the second time derivative becomes an -omega^2, so you end up with this equation with no time derivative, and that equation is called the Helmholtz equation. This quantity here is the wavenumber squared, so it is (2π/lambda^2).

To avoid writing hats for the next two hours, we are just going to drop the hat and remember that, in the Helmholtz equation, pressure is a complex quantity; it has the amplitude and phase built into it. I will even sometimes call the Helmholtz equation the wave equation.

The important point is that if we consider turbulence to be frozen in time and we consider lots of snapshots of turbulence, this equation describes everything, all the phenomena that we talked about, refraction, diffraction, scattering, geometrical spreading, everything is in this simple little equation.

The mathematics is not so important as having some insight into what is in there and what phenomena in the real atmosphere cause things to happen. The real atmosphere is three-dimensional but we are going to solve two-dimensional equations, and here is the justification for it.

[Transparency 31]

If you look at del in cylindrical coordinates, and you assume azimuthal symmetry, then this term here goes away. If you, further, define a little p that has cylindrical spreading built into it and plug it into the Helmholtz equation, you get this.

This term right here, $4k^2/r^2$, if you look at it at one wavelength, it is 160, so already at one wavelength it is down by over a hundred. If you look at, say, three wavelengths, it down by a thousand. You do not have to go very far from the source before you are looking at what is essentially a two-dimensional equation. That is often called the farfield Helmholtz equation.

Since I am really going to be talking about two dimensions, I am not going to use r very much, but will mainly use x, so we will be talking about this equation right here. It is a two-dimensional equation and I can assure you at this point that it is okay, because we have done 3D calculations and as far as mean levels go, the 2D works fine.

So that is a Helmholtz equation, but we are not talking about a Helmholtz equation, we are talking about a parabolic equation. So what is the difference?

[Transparency 32]

The difference is, if we look at a one-dimensional Helmholtz equation, what we see is that we have a second derivative with respect to x. That means we have, for this kind of time dependence, e^{ikx} and e^{-ikx}, which are the solutions. The first one goes to the right, the second one goes to the left.

If you want to extrapolate fields, you need a range derivative. There were a lot of people in the early days who just pooh-poohed this approach and said, "Look, I know how to solve equations with two range derivatives, why do you go through all this stuff? I'll just use blah-blah-blah's method and we'll just solve this equation directly."

The problem with that is similar to Phil's problem of backing up waves. When you deal with this equation, the back-going waves are just lying there waiting for you to make a mistake, however tiny. The numerical noise starts exciting these backward-going waves and as you try to move out you start getting crazy things.

The people who promoted that approach are not heard from very much any more. The approach that is taken is to split the two-way wave equation into two pieces, a forward-going equation and a backward-going equation. If k is a constant and you do all these indicated operations, you get back to the Helmholtz equation, but it is pretty clear that if I have a solution to this (parabolic) equation, it is also going to satisfy that (Helmholtz) equation. If one of these two terms is zero, it works.

This is a one-way wave equation, it has one range derivative, the plus sign is right-going, the minus sign is left-going and they satisfy the Helmholtz equation but they are one-way solutions. You cannot make an echo with them, so no matter how short the word length in your computer is, the wave always goes in one direction. That is very important if you want to get answers that make any sense.

This splitting of the two-way wave equations into one-way wave equations is the way people think about it nowadays. Tappert did not do it that way but he got a useful answer anyway. I will try to put the parabolic equation in some perspective using the viewpoint generally taken today.

Things get a little more ethereal and abstract when you go beyond one dimension. We split this one-dimensional Helmholtz equation; instead of k^2 we had k, so we think of k as the square root of k^2. But what happens when we try to do that with a two-dimensional wave equation? Instead of just k^2, we also have a second z derivative.

[Transparency 33]

Now this thing is no longer a number but it is an operator I call Q. By sort of faking myself out, I can make it look like a one-dimensional equation. If I just sort of blindly follow the mathematics, instead of the square root of k^2, we have the square root of Q. We can break it into a right-going and a left-going -- formally right-going and left-going -- wave.

We have this thing called the square root of Q. I cannot say that is 13 or 6, or something, it is an operator. What does it mean? To sneak up on it, what we can do is say, okay, let's take k to be our reference wavenumber, k_0. This k_0^2 here is just a number.

If you are talking about constant wavenumbers, you are talking about representing things in terms of plane waves. When I operate on a plane wave with a second z derivative, I get minus the vertical wavenumber squared. Up here, in the first equation, this is minus the vertical wavenumber squared, and we know that the vertical wavenumber squared plus the horizontal wavenumber squared is the whole wavenumber squared.

That tells me that $k_0^2 - k_z^2$ is the horizontal wavenumber squared, so this quantity is the horizontal wavenumber squared and the square root of that is the horizontal wavenumber. That is pretty obvious when you have k is equal to some constant number.

But what happens, then, if you have some general wavenumber that is a function of z and x and so on? This, then, is a true operator. Instead of numbers we have to talk about the eigenvalues of the operator. The eigenvalues of Q are the horizontal wavenumbers squared and the eigenvalues of the square root of Q are plus or minus the horizontal wavenumber.

How you approximate the square root of Q has generated a cottage industry among mathematicians. There have just been a lot of approaches. I am going to take the simplest possible approach; that is, to make a linear expansion. There are much better ways to do it but for our purposes a linear expansion is good enough.

[Transparency 34]

I have a viewgraph on what Tappert did but I think I will save it for the end in case there is interest and time in seeing what he did. It turns out, with Tappert's approach, you get the same answer but it does not tell you what you are doing, so you do not know how to do any better.

We wanted to make a linear approximation to the square root of Q and we had this equation with outgoing waves, we had the square root of Q, and there is what Q is. I am going to go back to saying let's think about the wavenumber here being a constant. When we do that, we get the same thing in pictures that we had before in symbols, the wavenumber squared is the sum of the squares of the x and the z components, so the horizontal wavenumber is given by that expression there. This is the total wavenumber. Take off the vertical wavenumber and it leaves the horizontal wavenumber squared. In general, when k is not a constant, these k_x's are the eigenvalues of the square root of Q; $(k_x)^2$ is the eigenvalue of Q, the square root, which is k_x, is eigenvalue of the square root of Q.

The idea is this, and the reason I drew this picture up here, is that if this angle is not too big, if your propagation is over angles that are not big -- what is not big? not big means less than, say, 10° with respect to horizontal -- in those cases, whatever the eigenvalue of this operator is, it is not going to be much different from k_0. You can see that k_x and k_0 are about the same.

What we can do, then, is linearly expand, and we do not expand about k but we expand about $(k_0)^2$. If you make that expansion of the square root of Q and just keep track of Q as a symbol that represents not a number but an operator, this is what you get for that linear expansion. If you plug back in the definition of Q, then you get that expression right there.

Now we have things that we have some warmth for, that is, a second z derivative, and numbers, and functions, so we do not have to be experts in Hilbert space, we can just be ordinary people.

That is the expression that I will be using to represent the square root of Q and, further, I am going to call this -- this is $(k+k_0)(k - k_0)$and the $k+k_0$ I am going to call $2k_0$ and divide by $2k_0$ so we will approximate that with $k - k_0$. Then the equation we end up with that is going to approximate the one-way propagation in the Helmholtz equation is given there.

I will not go into it right now, because I want to make sure we do not run out of time but, if possible, I will show you what Tappert did. Tappert ends up with essentially this result. He does it in a very straightforward quick-and-dirty way and he gets this equation right here.

[Transparency 35]

An outline of Tappert's original approach is given in this transparency. The pressure field, p, is written as a carrier wave times a modulating amplitude. The idea is that the modulator varies spatially much more slowly than the carrier wave. The expression for p is substituted into the Helmholtz equation. The carrier wave drops out, leaving an equation for the modulating amplitude, ψ. Since ψ varies slowly in space we can drop $\partial^2\psi/\partial x^2$. [Note: to show $\partial^2/\partial x^2$ can be dropped, consider a plane wave making an angle, θ, with respect to horizontal, where $\theta \cong 10°$.] Once $\partial^2\psi/\partial x^2$ has been omitted we have a differential equation that is first order in x. The equation is identical to that at the bottom of transparency 33, which was obtained by linearly expanding $\sqrt{Q}$, where $Q = \partial^2/\partial z^2 + k2$.

What he apparently did not know and what was discovered through work done in the seismic community is what he had really done was make a linear expansion of square root of Q. There are many, many other ways of approximating the square root of Q. An exploration geophysicist named Jon Claerbout at Stanford came up with a rational approximation. Mathematicians said, "Oh, we know what that is, it's a Padé approximation." Now Padé expansions, it's easy forget what you are doing physically. Using a linear expansion will keep us on track physically and it is perfectly good for our purposes.

DR. SPARROW: At the bottom of that page, is there a misprint? The second derivative with respect to z, +k?

DR. GILBERT: Right, you had up here a $-k_0$ that cancels that k_0. That is right. It scares me that people are checking so carefully. (laughter)

That last equation that I wrote really is kind of hard to make any sense out of, so I am going to do a little backstepping to get it into a form that is more physically interpretable.

[Transparency 36]

What I want to do is note that if you expand this ($\sqrt{k_0^2 + \partial^2/\partial z^2}$) linearly, you will get a $-\frac{1}{2}k_z^2/2k_0$. If you remember, k_z^2 is what the second z derivative operator gives you when you are operating on a plane wave, it just brings down ik twice. What that tells us is that this term right here ($-\frac{1}{2}k_z^2\ \partial^2/\partial z^2$) can be replaced with this ($\sqrt{k_0^2 + \partial^2/\partial z^2}\ k_0$) If I linearly expand this term, assuming this vertical wavenumber squared is small compared to k_0, and that is true because of small-angle propagation, and I put this back into the equation we were looking at, we suddenly get something that has direct physical interpretation.

The quantity, $\sqrt{k_0^2 + \partial^2/\partial z^2}$, gives the horizontal wavenumber for free-space propagation of plane waves. It is the square root of the total wavenumber squared minus the vertical wavenumber squared, which is the horizontal wavenumber squared. This term right here ($k\ \ k_0$) is a measure of the departure of the wavenumber from the reference wavenumber. What we are getting, then, is this term, $k-k_0$, which represents refraction. It will end up in an exponential and it will change the phase because k is changing relative to k_0. That gives the refraction part.

This part right here, $\sqrt{k_0^2 + \partial^2/\partial z^2}$, is nothing but the horizontal wavenumber for a plane wave in free space and it just takes care of diffraction. If you notice, Anthony talked about a lot of this stuff and the only difference here is I have this refraction term whereas in all of the Fourier propagation analysis that was talked about earlier there was no refraction.

The gist of this method is simpleminded and it is probably why it works: It is simpleminded and it works. If you do the split-step Fourier method that we are going to talk about, and your range steps are small enough, then we can do diffraction and refraction in sequence instead of simultaneously, so we will propagate in free space and then make a phase correction and then do it over and over, again.

It is basically saying over small enough range steps, I do not care that the rays are bent a little bit, I can make straight-line approximations and then just correct for the phase over that straight line. I will show you some of the mathematics that go behind that.

[Transparency 37]

Now let's look at the formal way of doing this. The reason I want to do this is because it brings in the notion of a "phase screen." There is the equation we were just looking at, this ($k - k_0$) being the refractive term, this ($\sqrt{k_0^2 + \partial^2/\partial z^2}$) being the diffractive term, and we want to define an operator A for the refraction and operator B for the diffraction, and the sum of the two we will call U.

Our equation becomes pretty simple-looking, a partial of P with respect to x is equal to some operator operating on P. The formal solution to this is to integrate U from 0 to x and put it in an exponential. If you plug that back in, you will find it does indeed satisfy the equation.

But now we have done something that really simplifies life as we have integrated over x. I will call this integrated value of U, U bar times delta x. Similarly, that means U bar times delta x is A bar plus B bar times delta x. This quantity where we have integrated over the departure from a reference wavenumber over x was a function in general of x and z, but now that we have integrated over x, it is just a function of z. This quantity is called a phase screen, so it takes something maybe this big and collapses it down to something of zero width but it has a vertical phase variation.

So A bar, which means the wavenumber integrated over x minus the reference wavenumber or the index of refraction integrated over x minus one -- here is what we mean by this n bar -- if we are talking about the turbulent part, there is some x variation. If we are talking about just the mean part, there is only z variation. For the mean part this just comes out to be delta x and cancels with this delta x.

Similarly, the B operator has no x dependence, so B bar is just B. If I look at the mean of n, the n bar is just the value itself, but for turbulence, where there is x dependence, we are integrating out the x variation and getting something that just varies with z.

Our formal solution looks like $\exp(\Delta x\ (\bar{A}+\bar{B})]$ operating on the field at x_0 and this carries the field forward an amount delta x. This x here is x_0 plus delta x. That advances the field, but in order to do refraction and diffraction sequentially we need to make an assumption. This assumption is that these operators commute, $\bar{A}\bar{B}=\bar{B}\bar{A}$.

Is that true? No, it is never true. So how do we get away with it? The reason we get away with it is that delta x is small enough that the higher order terms in the expansion of that exponential will get beat down. That lets us get away with it. The magic of this method is that small enough can be 100 wavelengths, so this method is semi-analytical. It has the sinusoidal variation built it. When we first did our parabolic equation calculations before about 1990, we just integrated the equations directly.

How big a range step could we take in those days? About a quarter of a wavelength. We went from a quarter of a wavelength to 100 wavelengths, so we were really happy. Things have improved a little bit in direct integration, you can now get a couple of wavelengths with some very fancy Padé expansions of the operators, but this method for outdoor sound propagation is, to me, the most physical and the fastest and the simplest way to go. There are a lot of positives for the method.

That is pretty much what we use nowadays. It is fast enough that we can go to three dimensions, we can do lots of statistics, so there is really hardly any reason for us to be looking at other things, but our competitors are always improving the other ways and coming up with stuff, but I do not care. (laughter)

[Transparency 38]

Here is the mathematics. There were lots of typos in the paper that somehow got corrected. These are not typos here, these are "pencilos." They are mistakes I made and I tried to correct them, but it does not matter. Just forget all that stuff. The only part I want you to pay attention to are the cartoons. The field shown at x=0 is a Gaussian. That is sort of a fat representation of a delta function.

The reason is, we do not want this to be a delta function, because when we Fourier transform it, the wavenumbers go on forever. The vertical wavenumbers go on to infinity. That corresponds to evanescent waves and very high-angle propagation. This representatio of a dealta function is fat enough so that when we Fourier transform it, the wavenumber spread in vertical is finite but is large enough for the calculation and is easy to deal with.

What all this mathematics is saying is this: You take the starting field and you Fourier transform it, and that is $\tilde{p}(k_z)$ instead of p(z). What does this mean? It means the amplitude of the plane wave component that has vertical wavenumber k_z. This represents the amplitudes of my plane wave decomposition.

Once you have done that decomposition, all you have to do is multiply this component by $e^{i\Delta x}\sqrt{k_0^2-k_z^2}$. That is, I have to multiply $\tilde{p}$ by exp[I(horizontal wavenumber times delta x] and that simple operation brings this Fourier amplitude down range in amount delta x, a very easy thing to do.

To continue, we have really got to get back to z-space, so we do an inverse transform. This is all at the same range here, even though I have spread it out so you can see it. We get this thing advanced in k-space and then we Fourier transform and bring it back to z-space. Are we done? Not quite.

There was stuff going on in here, there was refraction and we ignored it, so what are we going to do? We have got the thing back in z-space and this is just the phase screen, the integrated wavenumber, essentially, that accounts for all the phase change due to refraction in this region, so we just correct. We just need to multiply by the phase factor, $e^{i\varphi_1(z)}$.

In k-space the operator, $\exp\left(i\sqrt{k_0^2+\partial^2/\partial z^2}\,\Delta x\right)$, is simple. It is a multiplication. In z-space the operator, $e^{i\varphi(z)}$, is simple. It is just a multiplication. So what have we done? We have said z-space, k-space, free-space, propagation, back to z-space, correct for the phase, and now we have got the field advanced a distance delta x. So all three of these things are really right here on top of each other.

Then, you dot, dot, dot. That means at some point you have gone out as far as you want to go, say n steps, and you are doing the same thing on the nth step, you are bringing the Fourier amplitude forward, converting back to z-space, and making a phase correction, so now you have taken the field from zero to n delta x.

DR. MAYNARD: What happened to the z derivative, the second-order z derivative that was in those?

DR. GILBERT: The second-order z derivative, when you Fourier transform, becomes $-kz^2$. That is the reason you go to this Fourier representation, so that you do not have to do derivatives.

DR. MAYNARD: So that is where the B operator is put in?

DR. GILBERT: Yes, this is the B operator right here. You take your field, you let it diffract like it is in free space, but you know there are some phase changes in there, so you bring it back to z-space and make those phase corrections. You are basically saying the curved rays are approximated with straight rays, that is all it is.

I hope that was not too painful. The problem in using this method for outdoor sound propagation is that there is a ground surface and in the direct numerical integration it is very easy to put in the impedance condition for the ground. But with the Fourier method we are using here, no one in the past knew how to treat the ground.

We had looked at another method that involved a spectral method for horizontal wavenumber and the hope was that we could get that into a vertical wavenumber instead of a horizontal wavenumber integration. My colleague, Xiao Di, who was finishing up his Ph.D. at Ole Miss that time and working for me, and I had all sorts of schemes for doing that.

Xiao Di would work on this thesis during the day and do these calculations late at night and about the time he was leaving in the morning I would be coming in and he would say, "Dr. Gilbert, it no work." (laughter)

We finally realized what was going on and formulated a very elegant method that accounts for the ground. It does some things that we really were not expecting to get that we did get.

[Transparency 39]

Here is the mathematics. You can more or less forget the mathematics, but what you want to notice is you have one, two, three, four things. This No. 1 is just the thing we have been talking about, the direct wave. This No. 2 is a new thing. This R is the complex reflection coefficient for the ground computed for a locally reacting impedance surface and it generates a specular reflection off the ground.

This third term is something entirely new. When you do some contour integrations you pick up poles out of the reflection coefficient and those residues correspond to a surface wave. It travels horizontally to the ground and decays vertically exponentially. It also decays exponentially with range.

So we have a complete description that has the direct wave, the specularly reflected wave, and the surface wave. This really made the Fourier approach just what you need for acoustics.

MR. APOSTOLOU: Can you still do 100 wavelength slices when you have these surface waves?

DR. GILBERT: Yes. In fact, this expression, if you consider the sound speed to be a constant, this is exact. It is an exact semi-analytic solution. I can go as far as I want as long as the sound speed is constant, but because we are having to make corrections for the fact that it is not constant, we have to take some finite range step.

This is so much nicer than a bunch of numbers and direct integration. I can look at this term (number 3) and see how big the surface wave is.

MR. SIMMONS: When you started, when you were in free space, I assume that the nature of the diffraction or refraction is due to the variation in the sound speed, the temperature and velocity fluctuations, so this last slide leads me to think can you apply this method of diffraction to, say, objects? Just the one wave propagation.

DR. GILBERT: In some cases, you can. If you are interested in forward scattering, then this will work. In fact, that is how we treat turbulence. If you are interested in backscattering, there is none. It is explicitly excluded.

You said something that indicates that you are confused. You said when you go to k-space, then you said something about refraction. When we go to k-space, there is no refraction. We throw it away and we let the diffraction occur as if the sound speed is constant. Once we have gotten to where we are going to go, we make a correction due to the fact that there are phase variations along the way.

It is diffraction in free space, get back to real space and make a phase correction and then do it, again. The difference between this and what Anthony was talking about is we have this phase correction and that is really the only difference.

MR. APOSTOLOU: These surface waves are sort of an artifact to complete the model or is it measurable --

DR. GILBERT: It is measurable.

MR. APOSTOLOU: -- by experimental techniques?

DR. GILBERT: Lots of people have measured it. It is there.

Really, it is a fact that if you have porous ground and if you get the right kind of porous ground, like snow, you can get a whopping surface wave. Some fellows have shot pistol shots over snow for decades measuring the surface wave, so this contains it in a more or less analytic form.

DR. WAXLER: Can you....backscattering by....

DR. GILBERT: People, particularly in the underwater acoustics community, have generated some very nice schemes for taking into account single scattering in the backward direction, and it is pretty weak, so single scattering is a pretty good approximation. Basically, at each range step they compute the backward-going wave and it gets all added up.

For atmospheric acoustics it is measurable if you have a sodar. That is the principle it works on, but these levels are tiny compared to what the outgoing wave is.

DR. HAMILTON: By surface wave, you do not mean like a Rayleigh wave, do you?

DR. GILBERT: No. This is a wave that decays exponentially with height and exponentially with range. If you just look at the wave equation and you start writing out what are possible solutions, if you have a constant angle-independent impedance, which is what locally reacting ground acts like, you find that there are not only these propagating waves but there is an exponentially decaying wave that satisfies the boundary condition.

It is kind of like an evanescent wave. You see these propagating waves but you also discover there is an evanescent wave that also satisfies the equation. If anyone has a better explanation of surface waves, I would like to hear it.

MR. PETCULESCU: Does this surface wave radiate energy in the upper half?

DR. GILBERT: No, it is exponentially decaying. It is evanescent vertically. It just hugs the ground and you can measure it as a separate arrival.

MR. PETCULESCU: So it does not couble back into the upper half?

DR. GILBERT: In a flat range-independent environment it does not couple but if you put something in there, you get scattering out of the surface wave into other waves, and vice versa, scattering from other waves into the surface wave.

DR. GARRETT: This exists only if you have propagation into -- if you have a wave penetrating the surface, right?

DR. GILBERT: Yes, but the locally reacting boundary condition lets you ignore what is happening below, but you are right. What you actually have is this. I was going to talk about this, but it seemed too far afield.

[note: a hand-drawn transparency was used here. It showed an acoustic ray going from the air into the ground. The ray was refracted almost straight down.]

The porous ground has air in it but the sound speed in the porous ground is what? 50 m/sec?, very small. It is 350 or something up here in the air. What you get is whatever comes in is refracted straight down. No matter what angle it comes in, it goes straight down. Because there is nothing going on except refraction straight down, you really do not have to pay much attention to the physics.

DR. GARRETT: But that is the explanation for the evanescent waves. You have a fast medium on top of a slow medium.

DR. GILBERT: If you have a wave generated from down here, it will be evanescent up here, because this is slower than this, and that is one way to look at it. You can look at it as an evanescent wave coming from the other side. It is evanescent vertically. Then it will fall off exponentially and propagate like this.

DR. GARRETT: It is a total internal reflection of the wave that is in the ground

DR. GILBERT: Yes, that is a nice way to look at it.

DR. GARRETT: I just want to point out that that is exactly how insect eyes modulate the amount of light. They secrete a precipitate that scatters the evanescent wave.

DR. GILBERT: What Steve is saying is if I have a sound speed that looks like this and I have a field down in here in the ground, although it might be propagating to the right like this, above here you get total internal reflection, so it is going to have an exponential tail. Also, energy is getting soaked up like crazy down here, so it falls off exponentially, but you really do not have to think about that. I did not want to get into the ground. Jimmy is the dirt man and I am the air man. (laughter) ("It is soil.")

DR. GILBERT: Excuse me, soil. You can get more funding for soil than dirt.

DR. ARNOTT: On your equation there for the surface wave, that is propagating in the z direction?

DR. GILBERT: Yes, let me point out what is going on there.

[refer to Transparency 39]

DR. ARNOTT: Beta looks like it is dimensionless.

DR. GILBERT: Beta is 1/zhat and it is a complex number, zghat is a complex number, so 1/zhat is also a complex number. It appears up here -- right here -- so this is a complex number, beta is a complex number, and the z- part will decay exponentially. It will wiggle, but it will decay exponentially.

When you put it in here you find also you get an imaginary part that makes it decay horizontally. As the ground gets harder and harder, the imaginary part gets smaller and smaller and the decay takes longer and longer and it goes further and further, but this term out front that has beta in it, beta is getting to be very, very small, so you have this wave that goes forever with zero excitation. It is kind of a tradeoff.

DR. ARNOTT: Beta needs to be one over distance.

DR. GILBERT: You are right. I will have to check on that. Something is wrong there. I think probably what is wrong is that I should not have normalized it. Anyway, you are right, beta should have dimensions of one over distance. One great thing about this lecture is I had to remember what I did over the last 10 years, but you are right, there is something wrong there.

[Note: transparency #39 was corrected, $\beta = k_0/\hat{z}_g$ *not* $1/\hat{z}_g$*.]*

[Transparency 40]

That is basically all the introductions I have and I want to begin on the research story and how we used this. The very first work I didwas with Mike White. He wrote his Ph.D. thesis on using the parabolic equation in atmospheric acoustics and he finished up in about 1989, and left for a job -- he took Richard Raspet's old job -- and he was like most graduate students, he wanted to get out as fast as possible, so he left before we could put turbulence in.

My colleague, Xiao Di, Richard Raspet, and I carried on. What I want to show you is what Mike computed and why we knew something was wrong somewhere.

[Transparency 41]

These are Richard Raspet's data. You should appreciate what a wealth of information is in here. What we got is levels plotted versus frequency across the top versus range along the side. What are we plotting? What we are plotting is the result as a function of this parameter a, the refraction parameter.

Negative values correspond to upward refraction, positive values correspond to downward refraction. If you look at the lower frequencies, it looks pretty good. As someone said, at least everything is on the same page. But when you go over to higher frequencies, you are starting to see something that looks as if it is too low and if you go out in range, it is clearly too low -- clearly too low.

We could see that at the higher frequencies we were not even coming close to the data; we were like 80 dB below the data. It was pretty obvious that if we put some squiggles and wiggles into the sound-speed profile it was going to put some energy down there, but whether or not it would put anything that resembled reality or not was unknown to us.

We teamed up with Richard Raspet, who taught us something about atmospheric acoustics. We had the notion that we might be able to understand the data.

[Transparency 42]

The data looked like this. This is the relative sound-pressure level versus range and 0 dB means spherical spreading. These are levels relative to a spherical wave. This part is called the skywave, this is the boundary of the shadow, and this is the shadow region. If you do a no-turbulence prediction, it just goes down and down and down whereas the data are fairly flat.

[Transparency 43]

What I mean is this. This part that goes up in the sonified region we call the skywave. This angle here is small, 5° or 10°. If you are in this region, you are in the skywave part, it is basically spreading spherically. As you come across this way, you are crossing the shadow boundary and the levels start falling. When you get out in here, relative to spherical spreading, it is pretty much constant.

That is what the regions mean. You are up in the upward refracted part of the wave and you cross out of that, go across the shadow boundary, and you get out into the shadow and it is about -20, to -30 dB. If you just use mean profiles, you find out that the levels just drop precipitously, and that is not what the data do.

DR. WILEN: I do not understand that picture with the source. Are you doing sort of one ray?

DR. GILBERT: Yes. What we have is something like this. We have all sorts of rays and this is like the bounding ray. If a ray comes down at a steeper angle, it goes like this -- or like this. The ray that just barely skims the ground is the indication here. This is a representation of the shadow boundary in terms of rays. Of course, it is not really that sharp. Depending on the frequency, it can be very diffuse.

If we take a cut through here, we get spherical spreading, zero dB, then we cross the shadow boundary and it drops a lot, 20-30 dB, and then you come out into here and the levels are fairly constant.

These data were measured in the 1960s by Weiner and Keast. No one had ever explained this. This characteristic was in all the data and there was no explanation for it. So we said let's take a shot at that. We will take the code that Mike White wrote and we will put turbulence in it. It turned out that was an easy thing to do.

[Transparency 44]

How did we do that? Gilles Daigle and his colleagues at the National Research Council in Canada had measured temperature and wind fluctuations. They had fit the data, the autocorrelation function, with a Gaussian, because that was convenient for what they wanted to do and it looked like a pretty good fit to the autocorrelation function.

The trouble is, if you have a Gaussian in space and you Fourier transform the Gaussian, what do you get? A Gaussian. In fact, if you look at Gilles' data, the Fourier amplitude spectrum is not Gaussian at all, but he was not interested in doing what we were doing. He was interested in some forward scattering with no refraction.

The rest of the world said, well, we will blame the turbulence on Gilles and we took it as gospel. That way we were following what we thought was the purist approach, the scientific approach. We will take what is in nature and we will apply it and see if we get something that looks like what is measured acoustically.

The way we did the calculation was to take measured spectral amplitude for turbulence. First, you take the square root of it, and multiply that by a random phase. Then add up all the spectral components. What does it do? It generates fake turbulence. This is horizontal range this way and height this way, we are doing a 2D calculation.

Here is a meter, so you can see these wiggles that were generated by the Gaussian turbulence spectrum that came from Giles and his colleagues produces realizations of turbulence. This is what we have to crank into the P.E. calculation. In actuality, this was missing a lot of the small-scale structure, but it worked so well we felt pretty good about it, and the referee called it "pioneering work," a breakthrough in atmospheric acoustics. I have saved that somewhere. (laughter)

The first referee, I should say, on the paper Mike White and I put in, said, "So what? Why is this important?" We had to write a paragraph on why it was important.

[Transparency 45]

What we have here are some calculations with the parabolic equation. At this point we were just directly integrating numerically. If you look at the case -- this is crosswind propagation, so you are basically looking at the temperature. This is range. This is height. This bar over here gives you the scale, the relative sound-pressure level. Zero dB is spherical spreading and we had only so many shades of black, so we had to quit down here.

This is the calculation for crosswind propagation, weak upward refraction, without turbulence. We put into the index of refraction these little wiggles. I must remind you that these little wiggles -- look at this scale over here, 5×10^{-3}, so we are talking about a few parts in a thousand fluctuation in the index of refraction.

When you are looking at relatively small propagation angles, this is a small angle here. This is about 70/1500, so it 7/150, so what does that come out to be? It is about 0.05, so 0.05 x 60° is a few degrees; this is a few-degree angle.

This angle over here is going to be maybe 6°, so this is a small angle, and that one-part-in-a-thousand fluctuation in the index of refraction just pours energy down into the shadow zone. The question is, what are the numbers? We have lots of good data; what are the numbers?

What we do is we take a cut through here that represents the level versus range and that is what is measured, and we make a comparison to the measured data. These were data measured by Weiner and Keast in about 1963.

[Transparency 46]

And lo and behold, it worked sensationally well. The data are averaged -- these wiggles represent a realization. We do a trial. We create some fake turbulence with the right spectrum, with what we thought was the right spectrum, we propagate through it, and we get these wiggles.

This is a first trial.

This is a second trial with totally different-in-detail turbulence and we see that the average level is about the same. We did not want to make this too complicated. The data are the little black dots. The dotted line is the calculation without turbulence.

This is the crosswind, weak upward refraction.

This is the upwind, stronger upward refraction. We seem to be overshooting a little bit.

We go up to a higher frequency and things look pretty good for the crosswind propagation. Now we start to see something that was a little disturbing, that we are considerably and obviously low.

If I take this and put it up here, and this, and put it down here, for the calculation, it fits perfectly. I must have checked those figures at least a hundred times but I always concluded this is the right way it should go. So everything worked really well, except in this case, and we had some arguments about multiple scattering, and so on, but it turned out there was something more seriously wrong.

[Transparency 47]

Once we started using our Fourier method, we could do many, many realizations really, really fast and actually compute a mean level with standard deviations about it. It does not really tell you anything new, except you just get a smooth line. For the case where it did not work, you get a smooth line below the data, so averaging did not really change anything; what you see with your eyeball is what is there.

But it was nice that we had a really fast computation. We just brute forced through this. You generate turbulence, take a random number generator for the phases and the square root of the amplitude, then you add it all up and you get a bunch of stuff. You do that again and again and then you can produce by brute force all the statistics you could ever want.

We will leave Gilbert, Raspet, and Di stoutly defending their calculation against all complaints from my friends at the École Centrale, namely, Daniel Juvé, and everybody else in the world who knew anything about turbulence. They knew that the spectrum is not Gaussian, but we said we do not care, hey, look, it works, what is your problem? So I will leave us there and pick up in 10 minutes on the research story.

DR. GILBERT: Before I pick up again on the research story, there is something I wanted to mention having to do with inventing your own atmosphere. One of the things that people had used a lot for the mean-sound-speed profile was a linear profile. That was because you can write solutions in terms of airy functions.

But when we use a linear profile, it is not physical. The physical wind profile is logarithmic. When we used the linear profile we could not even come close to the data, so when we went back to what nature had in it, which is a logarithmic wind profile, it worked much better, the reason being, with a linear profile, you have a hurricane after a thousand meters, but this logarithmic thing kind of straightens out, so most of the refraction is near the ground and not so much up high. That worked.

We thought we were staying true to nature and using measured atmospheric inputs. It is not anybody's fault except ours. We lifted some work for the turbulence profile that really was not intended for what we wanted to do, but it worked so well that in our last episode we had me and Richard Raspet and Xiao Di stoutly defending our use of the Gaussian profile, and everybody in the world who knew something about turbulence knew that that was not what turbulence really was.

The people who came to our rescue were the NRC guys, David Havelock and Gilles Daigle's group. We got together and said, look, let's do an experiment with lots of frequencies and from that we will be able to infer from the acoustics what the proper turbulence model should be.

[Transparency 48]

Here is the Havelock experiment. It was at a glider port outside of Ottawa. We had 700 meters of runway. we propagated directly against the wind and so we had pretty good refraction upward. We had frequencies going from 40 Hz up to 940 Hz.

We were beginning to realize that we were becoming somewhat of a joke in the community, because everybody knew that the turbulence spectrum, if anything, was Kolmogorov and not Gaussian. So instead of trying to do what we were going to do originally, which was prove something that everybody already knew, we decided to go the reverse route.

[Transparency 49]

We had a little propeller anemometer. It sampled only once a second, so it did not give us a whole lot of wavenumber resolution but it gave us enough wavenumber resolution that we could see the beginning of the inertial sub-range. This is this cascade region from the large-scale stuff down to the tiny stuff that follows a universal slope (not necessarily the amplitude).

By getting this beginning of the Kolmogorov spectrum we could extrapolate down; that gave us the right level. That is what everybody in the world knew it should be, but our spectrum looked something like this and it fit in this region of around 500 Hz.

[Transparency 50]

We created realizations just as we did with the Gaussian spectrum and here is a picture of the upper end of the frequency. We have the wind going this way. We have the skywave and the scattering by turbulence down into the shadow zone.

[return to Transparency 48]

We had a whole bunch of frequencies. If you noticed, one nice thing about this experiment was that there was a whole bunch of microphones on the ground -- I do not know if that is the right number, but there was a whole bunch, because we calibrated them twice a day and it took about an hour.

[Transparency 51]

We looked at the levels along the ground and here you are in the sonified region, you pass into the shadow zone. We had this over a bunch of frequencies. This is 40 Hz. We had nine frequencies. This is six of them. This is the relative sound-pressure level versus range, 0 dB is spherical spreading, the range went out to 700 m.

At the lower frequencies you are getting lots of diffraction, so you really do not see turbulence, you do not see the edge of the shadow zone. If you kept going out farther and farther, you would. By the time you get to a couple hundred hertz, you start seeing things drop at the edge of the shadow zone and then become flat in that characteristic step pattern.

This was the lowest frequency and this was the highest frequency. This region is basically diffraction and this region, where it gets flat, is scattering from turbulence. That looked very good, it was a measured spectrum.

We wanted to say, then, okay, now let's look at this, compare the measured spectrum to the Gaussian spectrum.

[Transparency 52]

Here are some other frequencies with the measured spectrum. I am basically showing the envelope from 380 Hz to 940 Hz. That is about a little over two octaves. Here are the data. If you look carefully, you will see that, on average, it progresses downward monotonically as the frequency goes up, so this is the envelope calculated from a measured realistic spectrum.

[Transparency 53]

If we do exactly the same calculation with the Gaussian spectrum we had been using, this is what you get. At this point it became clear to us that using the Gaussian spectrum was the same mistake we had made before with the linear mean profile; we were inventing our own atmosphere and nature said, you know, "I've made an atmosphere for you, please use it."

When we did, it worked. When we invented our own atmosphere, it did not work.

DR. BASS: At what height did you measure the turbulence, because if it is a function of height it should be put in.

DR. GILBERT: I think they had a 10-m tower. The spectrum for the horizontal wind does not vary nearly as much as the spectrum for the vertical wind. The vertical wind changes dramatically with height. We looked at the JAPE data from a few meters up to about 30 m and I could not tell one spectrum from another; they all looked about the same for the horizontal wind. You have got to remember a factor of two or three in dB is not very much, so it was not that important.

At any rate, it is the slope, the shape of the spectrum, that was unphysical. We knew something was wrong and I am kind of a nitpicker. When I get something that looks right, I want to push it as far as I can. Either it becomes clear that it is really right or it becomes clear that it has limitations.

So we launched off on a project to really pick apart the calculation and look at what is physically going on. Earlier, about 1990 or 1991, we had done a perturbative calculation, so we pursued that route again as a way to understand the PE calculations. These brute-force PE calculations were giving us a result, but it is kind of hard to figure out physically what is going on. A standard procedure is plane wave perturbation theory, Borne approximations, which we had done before, but not with plane waves.

The difference is we did not use plane waves, we used the actual point-source wave without turbulence, so that was a distorted wave instead of a plane wave and the name for that procedure is distorted-wave Borne approximation, or DWBA. That involves an integral. You can integrate over space, real space, or you can integrate over wavenumber space, and we did it both ways. The first time we integrated over real space. The second time we wanted to know what part of k_z-space was important, so we integrated in wavenumber space.

[Transparency 54]

The thing I am calling the sampling function is the product of two Green's functions. I did not want to say that in public but that is what it is. Basically, this thing is being multiplied by the fluctuation in the index of refraction and integrated over space. A plot of this function at 210 Hz looks like this.

Green's functions are point-source solutions. Here, drawn by hand on the left, is one point-source solution from the source. You get a point-source solution for the receiver as well. You can think of a receiver like a source, except the rays go in the opposite direction. The region where these things cross is where you pick up the biggest contribution, so you can see this white region (from white to black it is 10 dB) is fairly well localized, so we are getting contributions from a fairly local region in space.

That is nice, but what we really wanted to know was how localized is the contribution in Bragg wavenumber space, the reason being that we had a notion that we were selecting out specific parts of the turbulence spectrum when we measured the levels in the shadow zone.

If, in fact, we were looking at a small wavenumber region, then we could understand it in terms of plane-wave Borne approximation, which assumes essentially that you have a single Bragg wavenumber. We knew we did not have a single Bragg wavenumber but when we look at the Fourier transform of this function and we look at it in Bragg wavenumber space, you see the contribution -- I do not remember what this range was, but it was a fixed range – the contribution is fairly well localized. It is saying the x component of Bragg wavenumber is essentially zero, there is some finite spread in the vertical Bragg wavenumber that is not zero.

The way to understand the result is if you have a scattering angle that goes like this (the angle is small with respect to horizontal), the horizontal component is hardly changing, but the vertical component -- this part -- is not zero, it is finite. In fact, because we do not have exactly plane waves, there is some spread over this scattering angle and that is represented by the spread over the Bragg wavenumber.

So in Bragg wavenumber space this was the sampling function and we began to see, then, gee whiz, we are really sampling over particular isolated parts of the turbulence spectrum.

[Transparency 55]

We assumed isotropic turbulence, so you can take everything from k_x and k_z into a single wavenumber, K, and do a one-dimensional integral over K. The sampling function then looks like this, it has a pretty long tail. When you multiply it by the spectrum, it kills off that tail and we see that we are sampling over a fairly small range of wavenumbers and from geometry you can compute the average Bragg wavenumber. We just have this blob up there and we take a point in the middle of it and we look at these angles and we compute the Bragg wavenumber and it falls right in the middle of that region.

What we did, then, is we went back and considered the frequencies we were looking at and mapped out the region in k-space that was being sampled and plotted that on the spectrum.

This is the Kolmogorov spectrum.

This is the Gaussian spectrum.

At 210 Hz we are sampling over this region of the turbulence spectrum. At 380 Hz we are sampling over this region. It is 600 Hz here and 940 Hz here. What you see, at 940 Hertz we should be here, but we are way down here somewhere. We said, aha, now we can start to understand the data and our calculations for the data.

[Transparency 56]

Let me continue with that story. The mental picture is that we have a scattering volume and although this is a bunch of plane waves, the spread in wavenumbers is not so big, so we can think of the average wavenumber coming in and, the converse, the average wavenumber coming out. There is a bunch of wavenumbers but they are not spread very big, so we can think of a plane wave in and a plane wave out. That is a very well-known mathematical procedure called Borne approximation.

[Transparency 57]

Here is what you get. I am not going to go into how you do it, other than it is an integral that you can do in real space or in wavenumber space. It says if I have a region of space with a specific spectrum for the fluctuations in the index of refraction and I have an incident plane wave and a scattered plane wave and I scatter over a specific angle, the magnitude of the Bragg wavenumber is 2K sin $\underline{\text{theta}_{scat}}/2$, and all that is, is the magnitude of the difference in the scattered wavenumber and the incident wavenumber.

What you find is that your scattered level, if I measure the scattered level down here, is proportional to the index of refraction spectrum evaluated at the Bragg wavenumber. (There are a bunch of multiplying constants, including the volume of the scattering.)

In a plane-wave sense you pick out a particular piece of the turbulence spectrum. In the actual experiment you pick out a narrow slice of wavenumbers in the turbulence spectrum.

[Transparency 58]

Now we can start to make some sense -- I repeat myself here. In plane-wave theory we are sampling at 210 Hz at -- I do not remember the range -- but we are sampling here and here. The Gaussian spectrum is a little too big here. It gets better here. It gets way too low here.

Let me jump back to the calculations that we did and say those same words for you.

[return to Transparencies 52 and 51]

When we have the real spectrum, the envelope of the calculation looks like this.

[return to Transparency 53]

When we have the fake spectrum, if I can line these up, it looks like this.

As you saw the Gaussian was too high at the lower frequencies and way too low at the higher frequencies. In effect, this is telling us that you have selected a piece of the turbulence spectrum and you guys invented your own atmosphere.

[return to Transparencies 52 and 51]

This is the real atmosphere.

[return to Transparency 53]

This is the Gaussian approximation.

The Gaussian is really okay as long as you are interested in only fitting a particular piece of the real spectrum, but because we went over such a large frequency region, it is impossible to fit the real spectrum over that frequency region and it shows up here, that we are way too high at 380 Hz and way too low at 390 Hz.

One other thing I want to mention briefly. Plane- wave theory is right, it explains lots of things, but it did not explain one thing. If you do plane-wave theory, you find that the frequency dependence is frequency to the one-third power and that is about 2 dB per octave. If you look at these data, it is about an 11-dB spread.

[return to Transparencies 52 and 51]

We said to ourselves that if what we were saying is right, why do we have so much frequency dependence? We should have only about 3-4 dB from here to here. But look, we have quite a big spread. (I do not know why I said 11, but that is the number I remember -- it does not look like 11 -- but it was more than we expected.)

[Transparency 59]

We said what is going on? What was going on is this. It was not the turbulence spectrum at all. It was the fact that we had those products of two Green's functions that defined the sampling in space. When you are at low frequencies, what you have is a dipole pattern. You have the source and its image in the ground and it creates a dipole pattern.

At low frequencies there is a big, huge, fat lobe as shown in the hand drawing on the upper right corner. If you have two fat lobes multiplied together, that is what the Green's functions do, you get this nice fat uniform region that you are collecting energy from. When you push up the frequency, instead of having one big fat lobe you start getting multiple lobes like this as shown on the hand drawing on the lower right.

When you overlap multiple lobes, you have zeros in there -- here and here -- and you have these three points of overlap and you have wiped out from the destructive interference, your sampling region goes from this down to this, and that is what, in fact, created the frequency dependence; it had nothing to do with the spectrum. That made us feel as if we were really on the right track here.

We are really picking this thing to pieces and pretty much that is how we left things, because we were all familiar with Borne approximation. But I do not want to leave you guys here. Borne approximation, if you do it all your life, it is a very physical thing, but if you have never seen it before, it is just like magic: He says that is true, why should I believe it?

[Transparency 60]

What I want to do is go back to the parabolic equation and help you to understand in a very direct physical way what is going on with these Bragg wavenumbers and how they sample the spectrum. What I want to give you is a "phase screen" explanation of scattering from turbulence.

[Transparency 61]

Here is the picture, a kind of review of what we were doing with the parabolic equation. I want to consider a plane wave coming in to this little slab of turbulence. If you remember, we are going to integrate over x. I am assuming that the mean sound speed is a constant, so all we have is turbulent fluctuations, and they are represented right here.

There is a plane-wave incident and this is the index of refraction fluctuations, but we want to integrate that over x and create a phase screen that just varies vertically. We bring this incident plane wave into – the z-axis at x=0. We have gotten there ignoring the fact that there was turbulence, and now, in order to go further, we have to make the phase correction by multiplying by the phase screen.

Then what happens is that you get the scattering, because we have this wiggly phase screen. To analyze this I want to assume this phase screen is made up of a sum of sinusoidal components with random phase, and some amplitude, S_j, which is proportional to the square root of the turbulence amplitude. This is, in fact, how we make fake turbulence. I should say, on a small scale, it is reasonable. On a large scale it is nonsense, because the structure in large-scale things, hundreds of meters, has nowhere near random phase, it is very coherent.

When you get down to meters, tens of meters, this makes pretty good sense, so in this sinusoidal sum that represents all these wiggles I want to pick out one component and I want to say this is the nth component, so we have some cosine modulation.

When a wave propagates through a sinusoidally modulated index of refraction it scatters waves primarily into two angles. The scattering occurs over plus or minus theta n, so you get just two scattering directions that is a direct function of what the wavelength is in this sinusoidal modulation.

What is happening is along these angles you are getting constructive interference. Whenever the difference between the scattered wave and the incident wave is equal to the vertical wavenumber in the atmosphere (this is all called the Bragg wavenumber), whenever your Bragg wavenumber is equal to the wavenumber in the atmosphere, then you get constructive interference and you can connect that to the physical angle this way.

If I set this equal to this, I can solve for this angle. If I take an acoustic wavelength of 1 m and the wavelength of this to be 5 m, we find out that the scattering angle of the constructive interference occurs at an angle of 11.5°.

I am not going to stop there. I am going to push this ad nauseam until you really have some physical feel for it.

[Transparency 62]

What does all this stuff mean that I just said qualitatively, what does it really mean physically? I am not going to make you look at the mathematics very much, except to say that all I am doing is applying the PE method, where we Fourier transform, then we propagate that transform as in free space, and then we multiply by a phase screen.

This already has a phase screen in it, so we have done that already and we are getting ready to do it again. This P at 0 as a function of z is an incident plane wave (this is x=0) multiplied by this phase screen, which has the integrated fluctuation that is now a function of just z times a wavenumber and times delta x.

I am going to assume delta x and delta n bar are small enough that I can expand this as one plus this argument. This argument, if you remember, you had one component of the phase screen that varied as a cosine, so we have one plus this cosine-varying term. When we do the integration and carry things into k-space, it gives me a spike at k_z incident.

What I have done here is I have written out the cosine in terms of its complex components and I have written out one as one. When we do this calculation, if you know the definition of a delta function, this is a delta function that says k_z is equal to k_z incident.

As a cartoon, if we plot the distribution of k_z, there is a big spike at the incident wavenumber, which is really the transmitted wave. At this level of approximation nothing happens to the transmitted wave; it actually just goes right on through. In reality, something does happen to it, but it is higher order. At this order of approximation, where we made this linear expansion, nothing happens.

This is the k-space representation, basically, of the incident wave. It had one wavenumber and it continues on with one wavenumber, so there is a big spike. Not so for this term that is modulated with a cosine. In fact, we get some different kinds of delta functions. We get the incident wavenumber plus the wavenumber for the atmosphere and the incident wavenumber minus the wavenumber for the atmosphere.

At this point things should be looking familiar. If I take a signal and I modulate it with another sinusoidal signal (i.e., multiply it by a different sinusoidal signal) -- you can think of it as something nearby to the actual signal, I am going to modulate it with a nearby frequency signal -- what do I get? The sum and the difference frequencies. And what do we call those things? Sidebands.

[Transparency 63]

We have this sinusoidal varying spatial modulator instead of a temporal sinusoidal and when we put the signal through that, it creates sidebands. Those sidebands are the wavenumbers for the constructive interference. If you think of turbulence as a superposition of a whole bunch of sinusoidal modulators, each wavelength spatially creates its own particular Bragg-angle scattering. That is the direction of constructive interference. It also can be thought of as the sidebands that are created by the modulation. So this is starting to look like familiar stuff.

That is basically it. The turbulence is creating to sum-and-difference wavenumbers instead of frequencies and those represent sidebands or they represent constructive interference.

Originally, our thought was we are going to let the acoustics tell us what the turbulence spectrum is. If we had pursued that approach, would it have worked?

Convert it back to z-space and you get an upgoing plane wave and a downgoing plane wave. What you are seeing, then, pictorially, is that you add or subtract the atmospheric wavenumber to the incident wavenumber. This is the sum-and-difference wavenumbers.

What you can say is the Bragg wavenumber is equal to the atmospheric wavenumber or you can say the scattered wavenumber is the incident wavenumber plus or minus the atmospheric wavenumber, so that is the same thing I just said.

DR. ATCHLEY: May I interject something? This is exactly what a diffraction grating does.

DR. GILBERT: Yes, that is what I am saying, it is a diffraction grating. I should have used those words, I usually do. The phase screen is a diffraction grating and if you think of it as a sum of sinusoidal components, it is linear acoustics, so you can add them up. You take one component of that. What does a diffraction grating do? It gives you wavenumber sidebands or gives you constructive interference in two symmetric directions -- it is symmetric with respect to the incident wave. Thank you, that is an important word. This phase screen is nothing but a complicated diffraction grating. You can analyze it one spatial wavelength at a time, which is what I just did.

Going back to what I wanted to say. The original thing we were going to do was experiment with lots of frequencies and we were going to infer what the spectrum should be. Could we have done that? Yes, I am going to show you that you could.

[Transparency 64]

But why didn't we do it that way? It got kind of embarrassing at this point to say we were going to do this fancy experiment to determine what everybody knew who owned an anemometer, that there was a Kolmogorov spectrum and that would be another "so what" paper. So we did it the other way, which is we said let's compare the Kolmogorov to the Gaussian. But could we have inferred the Kolmosorov spectrum? Yes, we could have. If you look at a given range in the shadow zone, what does it tell you? What it tells you is that the average magnitude of the scattered wave which is proportional to the amplitude of the incident wave and it is proportional to the magnitude of the Fourier component. If you account for the incident wave your measurement is proportional to the turbulence spectrum that is sampled.

Hence, at a particular angle you are picking out a particular Fourier component or a small band of Fourier components. That is pretty amazing; at a particular range you are sampling a tiny slice of the turbulence spectrum.

This Fourier component is proportional to what the turbulence spectrum is and it is evaluated at the Bragg angle and the Bragg angle is just the 2π/lambda times 2 times the sine of half the scattering angle. For small angles, the Bragg angle is $(2\pi/\lambda)$ times (scattering angle). What this tells us is the mean strength of the wave that is scattered at some angle, $(\theta_{scat} = \theta_n)$, is a measure of the mean strength of the nth component of the index of refraction, where that nth component corresponds to the nth wavenumber.

The conclusion is if we had gone that route, from the levels in the shadow zone we could have inferred the Kolmogorov spectrum directly from what we are measuring acoustically, but we did not go that route, because we knew what the answer was by that time.

DR. MARSTON: Back on your earlier page you indicated that you assumed a random phase, spatial phase, of the.... Can you give any information about that phase?

DR. GILBERT: The question is: When we generated fake turbulence we took an amplitude proportional to the square root of the spectrum and we multiplied it by a random phase and added them all up, so what happened to the random phase?

We are looking at mean levels. These levels get averaged over 5, 10, 15 minutes. Basically, with all these random phases, the cross products average out. If you look at the details of the time series, then those phases matter. In fact, no one to this point has made much use of the details of what is received as a function of time to deduce whether the random phases are really realistic or not.

My friend John Wyngaard says at the smaller scales random phases is an okay assumption. We are sort of inventing our own atmosphere again but we know in terms of mean levels those phases are not going to matter. If you are a signal processor and you care about these things, the millisecond-to-millisecond fluctuations, it is another issue.

MR. DEMIRCI: When I look at the graph there is a source and a receiver. I see the scattering from the scattering volume. Don't you also get the ground reflections?

DR. GILBERT: The question is, is the ground reflection included in the calculation? The answer is yes. This is a wave calculation. The only thing the parabolic equation neglects are the echoes coming back at you. In terms of forward propagation it has all the multi-paths, all the refraction, all the diffraction. It is a wave solution.

MR. DEMIRCI: Do you get the volume scattering at the ground scattering in the same manner? Does the ground scatter it differently from --

DR. GILBERT: If you remember, we had a kind of bounding ray. There is a certain ray limit that defines the edge of the shadow zone. Anything that hits the ground comes back up into this region. This is as far as you can go without hitting the ground. If I come in and hit the ground here, it is going to come back up here.

In the shadow region there is not a driving wave to create scattering. The only reason you can see the turbulent scattering here is because there is nothing else. It is like a night light in your room. If you leave it on during the day, you cannot see it, but when you turn off the overhead light and it gets dark, you can see it perfectly.

Turbulence scattering is all up here in the insonified region, it is everywhere, but up in here the direct wave is so powerful that the scattering by turbulence is just a negligible part of the whole field. Down here in the shadow region you have turned off the overhead light and now you can see the levels you measure are, in fact, above a few hundred hertz. What you are looking at is sound scattered by turbulence. That was basically what this story was about; we learned a lot about what is happening in the shadow region that you look at in the daytime.

PARTICIPANT: [Inaudible]

DR. GILBERT: The question was what about the ground bounces. The ground bounces never get into the shadow region, but the scattering from them is all there. If you remember, we talked about a dipole pattern. That is the direct plus the ground-reflected wave, how it creates lobes and those lobes define the region you are sampling from. At low frequency it is a big fat lobe. At higher frequencies it is a bunch of lobes and reduces the strength of the unscattered wave.

[Transparency 65]

Let me finish up now. This whole argument has been based on two-dimensional calculations and we felt confident, since it worked so well, that it was okay, but not confident enough that we wanted to leave it at that point, so we did some three-dimensional calculations. I am not going to go into much detail here.

[Transparency 66]

We created a pie slice and we had an impedance ground down here. Inside of this box there are mean profiles and there is turbulence. In order to enforce conservation of energy we did not want the turbulence scattering out and never coming back. We enforced periodic boundary conditions, so if turbulence gets scattered out this face, it comes back in over here in the other face, so we conserved energy.

Also, this angle, with present-day computers, is about 15°. Every time computer speeds double, this angle is going to double. Eventually we will have a 360° calculation.

[Transparency 67]

This was the mean profile. It is upward- refracting. I do not remember exactly what the turbulence model was. [Note: The turbulence model was Kolomogorov turbulence.]

Transparency 68]

Here are pictures you have seen already. This is the field as a function of range and height. The next one I will show you is cross-range, looking side to side instead of down range. What is interesting about this is you see 50 Hz, 100 Hz, and 200 Hz. What do you notice about the shadow region as you go from 200 Hz to 100 to 50 Hz? It goes away. Why is that? Why does it go away at low frequencies? Remember, this is a wave calculation. At 200 Hz there is a pretty clear boundary to the shadow region. As I go down in frequency, what is going to happen? What happens to edges when you go to finite frequencies? You get diffraction and the edge gets blurred out, so as I go down in frequency, the energy here is not coming from turbulence.

The energy here is coming from diffraction.

If you remember those curves, if you look at 50 Hz you do not see a shadow boundary. It is not until you get around 200 Hz that you have a visible edge to the shadow zone, and that is all these pictures are telling you, is that as you go from a higher frequency to a lower frequency you do not even have a shadow zone.

[Transparency 69]

Here is the impressive thing. This is cross-range. I am looking at the source down this way and this perpendicular direction is cross-range, and you see pretty much the same thing qualitatively, a nice clear shadow. As we go down in frequency, the shadow starts to fill up with diffraction. This is 200 m across this way and 200 m in height this way.

What we were interested in doing was investigating what it would buy you to put receivers up in the air. When you put an array of receivers and you want to find out what direction the sound is coming from, you have got to have coherent addition across the array.

Down in the shadow region in situations like this it is a mess. There is no coherence to speak of, so we computed the two-point correlation function. In other words, we said if we had a microphone here and a microphone here, how does the correlation vary as you separate the microphones?

[Transparency 70]

At low frequencies, and we had 0-, 25-, and 100-m heights as a function of separation. At 50 Hz you get good coherence out to 100 m separation, so you could have a very long array. Down on the ground at 100 Hz your correlation distance falls off very fast on the ground but at 25 and 100 m it is pretty good at 100 Hz. You go up to 200 Hz and what you find is that this one is in the shadow, these two are in the shadow, and the correlation distance between two microphones is very short, so long arrays will not help you. If you get up to 100 m you are back into the skywave and you get pretty good coherence again.

DR. BASS: Have you tried comparing those NRC data to see how close they are?

DR. GILBERT: We have compared it to our data. We did an experiment where we had two microphones and it is pretty reasonable.

What has NRC got?

DR. BASS: They did a bunch of different frequencies.

DR. GILBERT: This is different. This is different heights we are interested in.

DR. BASS: They were all about 1 or 2 m high.

DR. GILBERT: Yes, they get very short correlation lengths and that is what this is saying. When you get down to 0 m height, the correlation lengths in the shadow are very short, and it is controlled by what the scattering angle is (I will not go into that).

That is basically it and I just want to summarize what I talked about.

[Transparency 71]

I did not try to give you an overview of everything you can talk about in atmospheric sound propagation. I told you about a particular thing that I was involved in and in doing that I hope that I told you something about things in general.

I hope some of you found this interesting and will say, gee, this is not a dumb thing to study. If you are going to study atmospheric sound propagation, you need to know what the real atmosphere is and not invent your own atmosphere. We have done enough of that and gone down enough blind alleys.

You need to know the basics of propagation: diffraction, refraction, scattering, that sort of thing. The parabolic equation is not the be-all or end-all for everything in the world, we did a lot of our analysis using Borne approximation, but the parabolic equation is probably the most important tool right now in outdoor sound propagation and you need to understand what is going on with it if you are going to do research. There are other ways you can go but this is a workhorse method.

Finally, in order to not reinvent the wheel you need to learn what has been done in the past decade, because we have learned an awful lot about what is going on in the daytime. At night people have been able to get pretty good agreement with their data without turbulence, but that does not mean there are not some funny things going on.

If you want to make a lot of money as a noise- control engineer, you are going to be worrying about what happens when rock bands give concerts at night. They are not a problem in the daytime but at night – Xiao Di and I did a lot of calculations for Walt Disney. They were worried about their nighttime shows putting Mickey Mouse in people's back yards two miles away. They ended up, in the original Anaheim park, having to air-condition houses so they could keep the windows closed, and, in proposed parks they wanted to know ahead of time what was going to go on so they would not have to air-condition so many houses after the fact.

I hope there are no noise-control engineers here, but my impression is that noise-control engineering is in a very primitive state. No one even uses anything like this, so there are companies to be formed out there that do modern approaches to noise control. I do not know why I do not see it being done. If you are interested in starting your own business, this is a way to do it. There is plenty of business out there.

You need to get up to speed on what has been learned and, finally, there are probably other things, like learning how to use a microphone, a pre-amp, how to use a data recorder, I am sure that is something you need to learn, too, but that is not what I came here to talk about.

MR. WILSON: Would you say again how Kolmogorov came up with this, how he deduced the --

DR. GILBERT: He asked what spectrum would allow energy to flow uniformly through the scale size of eddies. He did that basically through dimensional arguments and the $k^{-11/3}$ spectrum is the only one that allows that to happen. That is what his argument was.

Later on, people measured these things and said, "He's right." But it does not tell you what happens on the very large scales. There is nothing universal there, it just depends on what the weather is, but once you get into this cascading region it is Kolmogorov.

Thank you very much for listening.

THERMOACOUSTICS

Robert M. Keolian
Pennsylvania State University

DR. KEOLIAN: Hello, everyone, good morning. Today I am going to tell you about something that I still find kind of surprising, which is that you can cool a six-pack with sound.

[Transparency 1]

I am going to be telling you about thermoacoustic devices. These are devices that use sound to pump heat, and you can use that heat-pumping in different ways. You can build a refrigerator with it, you can build a heat pump with it, or you can build a heat engine from it that will turn the heat into work. I will be defining all these things for you.

I want to say right off that the hard part about preparing this talk is in knowing what ink to leave in the bottle. This thermoacoustics business has gotten really big. There is a book out now by Greg Swift on the Web and I just recently taught a course from that. There is a bucket of math involved in this, and I could give you that, but I am not going to do that too much, because I have sat in on talks like this before and, before I was in the business, it took me a couple of years before I actually figured out how this thing worked. If I give you a lot of math, I do not think you are going to be able to take it away with you, so if you want to dig into it, the references I gave you earlier are a good way to dig in the literature, or this book from Swift.

What I am mostly going to do today is qualitatively describe how these devices work. We are going to start off with an introduction. There was a point I wanted to get to but had decided not to cover it because it takes an hour of math to get to that point. But Prof. Anthony Atchley did that math for us in his talk, so I am going to slip the point in, and that is the extra viewgraph page I passed out (Transparencies 8a - 8d). Also, another sheet of paper was passed out, which is a paper that a student of Prof. Steve Garrett's, Reh-Lin Chen, and Steve wrote on a demo, which I will be showing you.

As I said, I will be starting off with an introduction which will include a little bit of math, and then I am going to go qualitative and describe standing-wave devices and traveling-wave devices, the two classes of devices now being. Then I will conclude.

[Transparency 2]

Transparency 2 is a nice place to start if you want to do some reading. This is an article by Greg Swift—I am going to be mentioning him a lot. There is one more thing to mention: I have been stealing wholesale from my colleagues', students', and friends' viewgraphs and thoughts and I am going to rely most heavily on the work of Greg Swift, who is at Los Alamos.

This is the cover of a *Physics Today* review article on thermoacoustics, which is now a little bit outdated because it covers just this first class of standing wave thermoacoustic devices, but I show it because it is a nice color picture of sort of what these things are.

There is basically a pipe that is full of gas. That gas can be helium or air or some mixture of gases. There is a pair of heat exchangers. There is a speaker at one end and when you put sound into the pipe, entropy will be pumped along something between the heat exchangers called a secondary thermodynamic medium, a name given by John Wheatley, who got this business rolling way back when.

The guts of the thing are up at the top of this refrigerator in this diagram. Heat will leave the bottom heat exchanger and go into the top one. This bottom heat exchanger will get cold and his upper one will get hot—magically.

[Transparency 3]

Why would you want to do such a thing? There are no environmentally nasty refrigerants in thermoacoustic devices. We like to use helium, a mixture of helium and argon or helium and xenon or just plain air. These gases are usually under pressure, because more gas is better.

There are not many alternatives to the usual way that refrigeration is done, which is the process of vapor compression.

Thermoacoustic devices can be simply constructed, which means they are potentially cheap, and they are potentially very reliable. There is not much in them to break, so the maintenance on them should be pretty low. That is of interest to the Navy, for instance, who has been funding us for a while. They can be driven on waste heat, so there is sometimes no need for electricity to run these devices and so you can get free cooling. Those are the good things.

The bad things are that the efficiency so far has been kind of so-so. It is not bad but it is not great. It is a little bit less than what vapor compression is doing.

MS. POLIACHIK: There is a speaker in there. How do you run the speaker if you do not need electricity?

DR. KEOLIAN: There are two ways of doing it, with the speaker and without. With the speaker we are using electricity, without a speaker we are not using electricity. I will show you both.

[Transparency 4]

Let me start off with a little bit of thermodynamics just to define what we will be talking about. Let's talk first about what a refrigerator does for a living. What a refrigerator does is it pumps heat from a cold temperature up to a hot temperature. It is basically like a pump. To run that pump you need power coming in, which we call work, so the input to this thing is the work we are putting in, the output is the heat coming out the bologna sandwiches or six-packs and such.

The first law of thermodynamics, which is just energy conservation, says that Q_H, the rate of heat coming out of the hot side at absolute temperature T_H, is equal to the rate of work we are putting in, W, plus the rate of heat we are pulling out of the cold side, Q_C, at absolute temperature T_C. *[The symbols Q_C, Q_H and W should have dots on top of them to indicate that they are the time rate of change of the cold heat, hot heat and work, respectively. The dots were omitted to simplify the editing.]*

The second law—that is, the bogy-man entropy one—says that you cannot just move energy the way you might think you ought to be able to move it. Sometimes there are restrictions on it based on the randomness inherent in that energy. What it says is, this quantity, entropy, that Al Migliori and Jay Maynard talked about earlier, has to increase. The entropy leaving this engine—you have to take my word for it—is Q_H/T_H.

The rate at which entropy is coming into the refrigerator is Q_C/T_C. The amount leaving it has to be greater than the amount coming in, because any device can only create entropy or at least break even. This restriction is what the second law says.

We are going to define what is sometimes called the first law efficiency η_I. It is also called the "Coefficient of Performance," or *COP*. What this is, is "what you want" divided by "what you pay for." What you want in a refrigerator is cold bologna, so Q_C goes on top. What you pay for is the work coming in. That tells us what the first law efficiency is, Q_C/W.

If we are doing the very, very best we can (which we cannot do), then the entropy leaving the refrigerator is equal to the entropy coming into it, and there is no entropy production in the device. Way back when, Mr. Carnot and his dad figured out that there is a limiting thermodynamic efficiency, η_C (also called the "Carnot Coefficient of Performance," or *COPC*). To find it, you set the input and output entropies equal to each other. We want an expression in terms of Q_C and W, so you would go through the first and second law equations, eliminate Q_H, and you would find that Q_C/W is equal to $T_C/(T_H-T_C)$. Notice that this can be much greater than one, because that T_H-T_C can be a small number. That is the best you can do. So a refrigerator can pump more heat out of the cold stuff than you are putting in as work.

The second law efficiency η_{II} (sometimes called the "Coefficient of Performance Relative to Carnot," or *COPR*) is how well you did, η_I, divided by how well you could have done, η_C, and that number always has to be less than or equal to one.

[Transparency 5]

Now let's talk about a heat pump. Let's say you want to warm your house, so what you want is warmth. You could take electricity from the wall and just burn it up or you can take fuel and just burn it up in your house, and that is what people normally do, but you can do better than that.

You can imagine taking your refrigerator at home, sticking it out the door, opening the door of your refrigerator and aiming that outside. There are coils on the back of the refrigerator that normally reject heat behind the refrigerator. Put them on the inside of the house and try to refrigerate the outdoors. You are trying to make the outside cold, when it is already cold out there, pulling heat from the cold and putting it into your house.

So you pull in heat from the outside, put in work to run the refrigerator, but you get out both the work you put in, plus whatever energy you can pull from the outside, so you win—you get more heat out than the energy you put in as electrical work.

Now, the equations end up looking the same. The first law is the same as it was before. The second law is the same as it was before. All we are doing is redefining what is input and output here.

The first law efficiency is now Q_H/W, because that is what you want divided by what you pay for. To get the Carnot efficiency you take the first and second law equations, but this time get rid of the Q_C, and you end up with $T_H/(T_H-T_C)$. Notice that this can also be greater than one. The second law efficiency is η_I divided by η_C, as it was before, and that it is always less than one.

By the way, let me teach you a little trick for remembering these Carnot formulas. When no entropy is produced the work is proportional to T_H-T_C. The Q_C is proportional to T_C. The Q_H is proportional to T_H. I actually do not work the equations out any more. I memorized the trick. If I want Carnot efficiency for a heat pump, I say I want Q_H/W for the Carnot case. I know that Q_H is proportional to T_H, I know that W is proportional to T_H-T_C and I just write $T_H/(T_H-T_C)$ down. It is a handy thing to remember.

[Transparency 6]

Lastly, let's talk about prime movers. The word "engine" is used in two ways. Sometimes the it is used to include all these devices. More often "engine" is used to mean something that takes heat and turns it into work, like a steam engine, like a locomotive. "Prime mover" is another name for "engine" used in this sense.

Now our input is like the burning coal in a locomotive, or Q_H. The output is the work W, turning the wheels on the locomotive. Waste heat goes out the smokestack, which is Q_C. The first law is the same as it was before—Q_H has to be Q_C+W.

The second law says that the entropy leaving the engine has to be greater than or at best equal to the entropy coming in (meaning that in general there is some amount of entropy produced by the engine, which at best is zero), so Q_C/T_C and Q_H/T_H swap positions in the inequality $Q_C/T_C >= Q_H/T_H$.

The first law efficiency is what you want divided by what you pay for, so η_I is W divided by Q_H. The Carnot efficiency is when everything is working perfectly, so if you use the trick I just taught you—W is proportional to T_H-T_C, Q_H is proportional to T_H—you get $\eta_C = (T_H-T_C)/T_H$. If you look at that, it always has to be a little less than one. Notice that it gets bigger and bigger as T_H gets hotter and hotter. That is why people like really hot running heat engines, like the turbines on jet planes; they like to run them really hot.

The second law efficiency is the same as it was before, how well you did divided by how well you could have done.

MR. PORTER: So there must be a reason you move your refrigerator and stick it out your door and heat the house at the same time?

DR. KEOLIAN: Well, you can, kind of. There is sort of a reason. Though commercially you do not see it too much, I believe. The conventional vapor compression relies on a phase transition of the Freon refrigerant. The Freon goes from a liquid to a vapor phase, but there is only a limited temperature range over which that will happen, so conventional heat pumps do not work very well when it gets really cold outside, so people tend not to use them.

Plus, the work you are putting into the conventional heat pump is probably in the form of electricity. When you use electricity in your home, you have already suffered a lot of inefficiency getting the electricity from where it is generated to where your house is. Electricity is a little more expensive than the fuel would be if you run the heat pump off of gas or oil. I am going to be proposing that what we ought to be doing is making heat pumps running off of fuel; that is sort of how I am going to end this talk.

Steve, did you want to say something?

DR. GARRETT: Just that it is not a little more expensive, it is anywhere from 5 to 15 times more expensive to burn electricity than it is to burn gas. That is a lot.

DR. KEOLIAN: Is it? That is a lot.

[Transparency 7]

If Greg Swift ever reads these notes, here is a viewgraph I lifted right from him about length scales. There are various length scales in thermoacoustic engines that we should be aware of.

Along the propagation direction, which we are going to call x, we have the wavelength, which is the speed of sound divided by the frequency. In this business we tend to use a for the speed of sound instead of c, because we are trying to use c for a whole bunch of other things, like heat capacity and compliance.

There is also the gas displacement amplitude, which we are going to call x_1, as before, in the previous lectures. These are the acoustic quantities that we are talking about, and x_1 would be u_1, the particle linear velocity, divided by the angular frequency ω. Perpendicular to the acoustic direction we have two characteristic lengths that previous lecturers have been talking about, the thermal penetration depth, δ_K, and the viscous penetration depth, δ_ν..

I am going to be showing you these things in kind of gory detail with Greg's animations, so we will come back to these. The thermal penetration depth is the square root of twice the thermal conductivity divided by the angular frequency and $\rho\ C_p$—where C_p is the heat capacity at a constant pressure per unit mass, ρ is the mass per volume. When you multiply ρ and C_p together you get the heat capacity per unit volume, which is a nice way to think of that little combo.

The viscous penetration depth is the square root of twice the viscosity divided by ω and ρ. The two penetration depths are the distance over which heat or viscous shear forces can propagate from the walls in about an acoustic period. We will see how that happens with the animations. It turns out these dimensions are typically about a tenth of a millimeter.

They are both about the same, this δ_{v} and this δ_K, so that their ratio is about one, but the δ_K tends to be a little bit bigger than the δ_{v} and that is a good thing for us, because good things are going to happen within that δ_K and bad things are going to happen in that δ_{v}, so we want the δ_{v} to be as small as possible.

In regular audio acoustics, like I am yacking at you now, the particle displacements might be a very small fraction of a millimeter, maybe just a few micrometers. The penetration depths are about a tenth of a millimeter and the wavelength is long—it might be about a meter or so. But in thermoacoustic engines we kind of reverse this. We get pretty big displacement amplitudes. At Penn State we are building a refrigerator that is big enough to cool a house and in the middle of it the x_1 is a foot.

[Transparency 8]

Transparency 8 is a condensation of the theory of ordinary acoustics and what's different in thermoacoustics. In ordinary acoustics we do an order expansion for various quantities of interest. The pressure is a mean pressure p_m plus the real part of p_1, the acoustic part, which is a function of x times $e^{\wedge}(i\omega t)$; we'll always be assuming harmonic time-dependence.

There are four equations that describe the physics that is important in the propagation of sound. There is the momentum equation, which is basically $F = ma$. You can see it in there; dp/dx is basically F, $i\omega u$ is basically a, and ρ is basically m.

Then there is the continuity equation. The first term $i\omega\rho_1$ describes the rate at which density is changing in a little volume, and that is balanced by the second term, which describes to the rate at which stuff is coming in and out of the little volume.

Third, there is an equation of state that tells you what the gas does in response to the pressure swings or the density swings. It is just basically the ideal gas law.

But there is usually a cheat, you know. When deriving the speed of sound, normally no one ever says much about the fourth important equation that describes the movement of heat. The poor heat-transfer equation always gets neglected. Somehow, somewhere, somebody usually slips in that the sound is adiabatic without deriving it. It was done here in the equation of state when p_1/ρ_1 is equated to $dp/d\rho$ with the derivative to be taken at constant s—at constant entropy. The heat-transfer equation, which I did not write down here, is what tells you that you should do that. Well, in this business you have got to deal with that adiabatic or isothermal thing—or in-between—pretty carefully, and we will be doing that qualitatively later in the talk.

We then normally combine all these equations to get the well known wave equation.

However, a fellow named Nicholas Rott, about 30 years ago, carefully figured out what would happen if you put sound in a tube that had a temperature gradient along it. The tube could either be really big or it could be really, really tiny. We are going to allow heat to go between the gas and the tube walls. Because of the temperature gradient along the tube, the gas will move between hot and cold regions through the cycle. The temperature gradient and the heat transfer to the walls leads to Rott's more elaborate wave equation.

It is kind of like what we just had. He had to throw in that heat-transfer equation pretty carefully, and there are a bunch of hard steps to get this new wave equation. But by the time you are done, you end up with something that looks kind of the same.

In the normal wave equation the first term is p_1. In the new wave equation we have p_1 too, but loaded up with other stuff $1+(\gamma-1)f_\kappa$. Prof. Atchley described what this f_κ is and I am not going to go into a lot of detail about it, it is too hard, but there is stuff hiding in here. The f_κ describes the effects of heat going in and out of the wall, and how it dissipates some of the wave energy. If you make a really good thermal contact, if you have your pores really small, this f_κ is going to turn out to be equal to one (we will see that) and the wave is going to change from adiabatic to isothermal propagation.

In the second term, we have the second derivative, basically. There is one derivative, and next to it is the other, but there is all this other junk around it too. More f's—I hate these f's. Sorry to you guys from Ohio, but I really hate the f's, they are ugly when you write them out. These things have confused me for a long time, but they are the core of the game.

Then we get this third term and that is sort of where the guts of thermoacoustics is hiding. In this third term, notice it has a dT_m/dx in it, the mean temperature gradient. Gee, you look at that thing and ask, what do I do? These f's are constants, they are just complex numbers.

Here is how I look at this equation. I like the mass on the spring, I can handle that one. So I am going to say this equation in space is going to be like an equation in time. This middle term is kind of like d^2p/dt^2, if we think of space as being like time. That describes the mass part of the mass on a spring.

The first term has just some stuff in front of it with a p_1, so that is kind of like the *kx* term, the springy part. The last term has a single derivative in it, *dp/dx,* so it is sort of like the resistance term. But notice that I have got this dT_m/dx in all this junk in front of that resistance-like term and I can mess with that. I can make dT_m/dx positive or I can make it negative, so I can make this thing look like a positive resistor or a negative resistor, willy nilly, by how I move the temperatures around, so I can get something that grows or decays, this time in space rather than in time. That is the good stuff that is happening.

[Transparency 8b]

Here is the theory part I wanted not to give you on two or three viewgraphs, but Prof. Atchley kind of set you up for it and there is a nice little nugget in here that is worth getting at. You will be rewarded with a little demo if you stay awake. We'll start with transparency 8b and come back to transparency 8a in a bit.

Let's consider a little length of pipe. This could be a really little piece of pipe, it could be like a tiny little pore in something, or it could be a big fat pipe. We have a mean temperature, T_m, on one end, and T_m+dT_m on the other. We are going to have p_1 on one end, and p_1+dp_1 on the other end, and U_1 coming in and U_1+dU_1 coming out.

This is a capital U_1, by the way. What I was showing you before was a lower case u_1. The lower case u_1 is the usual meters-per-second linear velocity of a particle. Capital U_1 is the volume velocity, the time rate change of volume, or you can think of it as u_1 times the cross-sectional area A.

This little element has a length *dx*. What we are going to do, instead of that wave equation I just showed you—instead of combining those four equations I told you about, momentum, continuity, equation of state, heat-transfer equation, and getting rid of a bunch of variables and leaving just a second-order differential equation for just p_1—what we are going to do now is describe the sound with two first-order differential equations; an equation for dp_1 and an equation for dU_1.

Now these f's might make a little more physical sense if we can keep the two first order differential equations separated. We are going to figure out dp_1 and dU_1 as a function of p_1 and

U_1 and imagine starting at one end of our device and working ourselves across and figuring out the changes and all these things go and see where we end up on the other end.

The momentum equation, $F = ma$, is written here, and now perhaps it is in a form that will make a little more sense. Notice first that we have $i\omega U_1/A$; that's the acceleration. Then take that dx, stick it under the dp_1, so we have dp_1/dx is just about equal to ρ_m times the acceleration. That is $F = ma$, except that there is this $1\text{-}f_v$ in there. What the heck is that? This factor is taking into account the effect of the viscosity at the walls, through this funny f function. You can think of it as modifying the average ρ_m, making it complex. If there were no friction, then the mass in the tube would just be moving as a plug. But because there is friction at the walls, the distribution of mass motion is a little funny. There are phase shifts along the walls and it is as if the effective mass got changed. We are going to roll all that in to this screwy $1\text{-}f_v$ factor.

Next we are going to get rid of these f's in favor of something perhaps a little easier to understand. Let's beat on that equation and force it to be in real and imaginary parts. That f function is a complex number and there is an i sitting there, so let's force this equation into the second form shown, $(i\omega l\, dx + r_v\, dx)\, U_1$. We are going to put the reactive part in that $i\omega l$ term. It is going to be the mass, effectively, of that plug of fluid. Then we are going to lump all the dissipation due to the motion of the fluid into the r_v term.

Let's think of this as an equivalent circuit. We are going to have p_1 on one side, $p_1{+}dp_1$ on the other. This is an equation for that dp_1. It is kind of like the voltage drop, except it is not a voltage drop, it is more like voltage rise. You should think of this as pressure rise in order to get the signs right. In the voltage convention the value on the left would be higher than the one on the right. The convention in calculus and all the math you have ever done in waves is the other way, that the right-hand side is bigger than the left-hand side.

The expression for dp_1 is as if we have an inductor and a resistor in series. The inductor represents the imaginary part of the terms in front of U_1, the resistor represents the real part. Here are two equations, if you like, that give you the inductance per unit length and the resistance per unit length in terms of stuff you know and these screwy f's, real parts and one minuses and imaginary parts. It is a big mess. But we can get these f's, there are equations for them, so it is not a big deal.

The continuity equation is the next one. Continuity is dU_1 in terms of p_1 and U_1. Volume flows in, but some of the volume disappears because the gas compresses when it goes in our little tube, so there is a bit that is proportional to p_1 and that compressing part is described by the $i\omega C\ dx$ term. Part of the volume gets lost because heat is coming in and out of the walls and that is described by this $1/r_k$ term. The sound is changing between being adiabatic and isothermal at the walls a bit.

Then there is this weird thing, $e\ dx\ U_1$, which is going to occur because of the temperature gradient at the walls. If gas goes, say, from the cold to hot region, it picks up heat from the walls and expands, and so you have gotten more volume velocity than you had before—more volume.

We can describe these terms as an equivalent circuit, if you like equivalent circuits, in terms of three parts, a capacitor, a resistor, and this weird little current source that gives a current proportional to the current flowing through the big wire up on the top.

If you do not like equivalent circuits, all the stuff is in the equations; the equivalent circuits are there just to help you see the equations if that is useful to you, if you have a EE background and are used to circuits.

We have taken these f's, which I hate —

DR. GARRETT: How do you feel about the f's, Professor? (laughter)

DR. KEOLIAN: —and turn them into five numbers, which I kind of like, l, r_ν, compliance C, r_k, and this e.

DR. MATULA: I am a little confused. In the equivalent circuit, normally I would think of a resistor as being a real part and would not have a frequency dependence, yet the equation shows it is the imaginary part with a frequency dependence.

DR. KEOLIAN: There are frequency dependences in those f's. I cannot give you a fast answer to that.

DR. ATCHLEY: It is tied up because of the $i\omega$ is the first —

DR. KEOLIAN: Yes, that is why the real versus the imaginary part is switched.

DR. GABRIELSON: But it is true that the resistances do have frequency dependence. That is true. It is not like —

DR. KEOLIAN: Right, but it is not strictly proportional to that ω. In one regime it tends to go as the square root of ω. All these weird things, all these things you can think about as being a problem are all wrapped up in the f's and that was the hard thing that Nicholas Rott did 30 years ago, is figure all this stuff out and stick it in those f's.

DR. MIGLIORI: It is also true that everybody but EEs takes the dissipative part as imaginary. Every branch of science except —

DR. KEOLIAN: And acoustics, buckaroo. (laughter)

One more thing to drag you through here and then the reward. We can combine all this into a sort of combined equivalent circuit for the whole thing.

[Transparency 8c]

Here are our five parts: Inertance, sort of a mass per unit length; a dissipative part—this is describing pushing this viscous gas through the tube, it gives you loss when there is velocity through the tube; the capacitor describes the compressibility of the gas; the second resistor describes the thermal loss from the gas (that one is a little tricker to see and I will use the animations to help you through that one). These first two resistors are bad things, all they do is produce entropy and nobody likes entropy unless you are trying to warm up your house. Then there is the current source, and this turns out to be the good guy. It is a little freaky to understand, but the following may help.

At any point the pressure and the volume velocity are related by the acoustic impedance *Z*. So let's do that, let's let $U_1 = p_1/Z$. I am going to take that continuity equation that we had earlier and I am going to substitute e/Z with a p_1 for eU_1. Now the current source looks like a resistor. (It has an imaginary part too, but I'm going to ignore that.) I am going to call it r_{TA} and it is going to be e/Z, which is equal to this funny combination of these f's and the σ. The σ is called the Prandtl number.

I can write this thing now as a resistor. These first two resistors are bad resistors, this new one is a good resistor. I can control this good resistor, I can control the sign of it with dT_m/dx and also with *Z*.

The *Z* comes from our choice of how we put the acoustic resonator together. We can control the ratio of pressure to volume velocity. We can make a traveling wave, for instance, and have that *Z* be a real number. Or, we can create a standing wave and have that *Z* be an imaginary number. We can mess around with the standing wave ratio and make the phase of *Z* anywhere in-between. Our goal is going to be to make the conductance $1/r_{TA}$ associated with the good resistor to be big and real. The funny combination of f's and σ times dT_m/dx times $1/Z$ we want to make have a big real part. If we have a secondary thermodynamic medium—somebody hands us a wad of cotton and we want to make the best refrigerator you can out of it—what we can do, I believe, is figure out what complex value the term with the f's gives, and then adjust the Z to have the equal-and-opposite phase of it so that we can make the combination real. Then we can mess with the dT_m/dx to make a nice efficient device.

DR. MIGLIORI: What happens when the Prandtl number is 1 and nothing special so f_κ and f_ν become equal, is that what it is?

DR. KEOLIAN: That might be so, yes, that is probably so. I think that is so. I'll go back to transparency 8a to answer this.

[Transparency 8a]

Let me show you a plot of these f_ν's and this funny combination of factors here. That whole weird combination of factors here is shown by these graphs. The solid line is for this Prandtl number of 0.7. The dashed curve is for Prandtl number of 0.2. I really should have defined the Prandtl number for you.

The Prandtl number goes down for low-viscosity gases. The viscous penetration depth goes down for low-viscosity gases. The Prandtl number is the square of the viscous penetration depth divided by the thermal penetration depth.

Our student, Ray Wakeland, plotted these things out. He could not figure out what this whole combination of f's and σ looked like. A Prandtl number of 0.7 is appropriate for the noble gases, helium, argon, xenon, or for air. A σ of 0.2 is the lowest thing you can get with the right mix of helium and xenon. A low Prandtl number is a good thing. It means you have lots of good thermal effects, not too much viscosity. Viscosity is always bad.

But this funny combination ends up not depending too much on the Prandtl number because of the weird near cancellation between f_κ - f_ν in the numerator and $1-\sigma$ in the denominator. Both the $\sigma = 0.7$ and the $\sigma = 0.2$ curves end up looking like the dotted curve, which is just the raw f_κ.. That's pretty much the answer to Dr. Migliori's question.

Let's go back through our moral over here. The curves are plotted for parallel plates, that is the pore I have in mind. The curves are an average over the width of the cross-section of all these strange effects I have been yakking about, giving you a sort of effective mass, and an effective compressibility of the pore. It is all rolled into these messy f's.

We are going to talk about two types of media pretty much. The first is called a regenerator. It has really teeny pores compared to the thermal penetration depth. It is going to be used in the Stirling cycle, which is kind of like a traveling wave. The regenerator case is going to lie on the left-hand side of these curves, at small y_o/δ_κ. For both the f's and this whole screwy combination of f's and σ, the real part is going to be near one, the imaginary part is going to be near zero.

[Transparency 8c]

When that is so, then in the expression for $1/r_{TA}$ the funny combination of f's and σ is close to being one. So we want the impedance Z to be nearly real to get the biggest good stuff. That means we want something like a traveling wave when I have really small pores.

[Transparency 8a]

On the other hand, the traditional standing wave devices, which I will be describing in the next part of this talk, were developed first, historically. People figured out that they wanted to pick y_o/δ_κ to make the f and σ combination have the largest imaginary part.

[Transparency 8c]

Back to the expression for $1/r_{TA}$, if the combination of f's and σ has a large imaginary part, then to make $1/r_{TA}$ have a large real part we want Z to have a big imaginary part, more like standing wave.

There is, I believe, a whole continuum in-between. Given any hunk of junk that you want to make a thermoacoustic refrigerator out of, there is a standing-wave ratio that you want that will make $1/r_{TA}$ the best. This is the unifying perspective that I believe Greg Swift meant in the title of his book *Thermoacoustic Engines: A Unifying Perspective*. I think this is the unifying part. You can see both cases of thermoacoustic devices all in one fell swoop with these difficult equations.

[Transparency 8d]

All right, no more punishment. On to the reward. I am going to tell you what is so swell about a negative resistor. When that resistor is negative, we are going to generate sound from heat. That is a good thing. When that resistor is positive, we are going to generate refrigeration from sound. That is also a good thing. So that resistor does good things for us whether it is positive or negative; it is going to be giving us refrigeration or it is going to be giving us sound.

What I am going to be showing you here is an analog of that sort of thing with a tuning fork. I am going to show you that you can actively make a resistance which is what the thermoacoustics does. What I have is a tuning fork with a couple of magnets on it.

Let's start with the tine on the left. I have this funny yoke here with a pair of magnets and a coil of wire. The coil I took out of a relay. The yoke is oriented so that the magnets are just off the axis of this coil. When the tine moves in and out the magnets go in and out closer and farther from the coil, changing the magnetization in that coil.

The voltage generated by the coil is proportional to the rate of change of the magnetization, so the voltage I get here is proportional to the velocity of the tine. I am going to amplify that with a preamplifier. I am going to phase-shift it for a reason I will show you in a second, and then stick it back in on the other coil.

The other coil, on the right of the diagram, is aimed at the other tine. There is another little magnet there and the coil is going to generate a force on that magnet. That force is proportional to the current through the coil, unfortunately not the voltage through the coil. The amplifier gives me a voltage rather than a current.

The current lags behind the voltage in an inductor. This coil has a fair amount of inductance, so I am going to compensate for that by throwing in a phase shift of the appropriate sign. Part of the trick in the demo is to orient these two coils perpendicular to each other so I do not make a big transformer in-between them. [Demonstration]

What I want to show you is an active resistance. On this oscilloscope that is projected on the screen over here, the upper trace shows the velocity of the tines. The lower trace shows the current through the driving coil, and I have it disconnected from the power amplifier right now. I whack it. You can see the decay of the velocity signal on the oscilloscope.

Now I am going to hook the driving coil up to the power amplifier. I have the preamplifier on a low gain. Since I have a velocity signal from the left tine and I can apply a force to the right tine I am just going to hook one up to the other and apply a force to the tuning fork proportional to its own velocity.

As far as the tuning fork is concerned, the applied force looks like a viscosity. But I can mess around with the sign of it electrically by flipping the magnets or flipping a polarity switch on the preamplifier. So I can make that viscosity look positive or negative.

DR. SABATIER: Could you point to me on the diagram where the voltage and the current that are being discussed are?

DR. KEOLIAN: There is a little resistor in series with the drive coil. The bottom trace is measuring the voltage across that resistor, which is proportional to the force or the current in the coil.

The other side of the oscilloscope is measuring the voltage coming out of the preamplifier, which is proportional to the velocity. (If you look at this demo closely, I cheated, I did not bring my phase-shifter with me. I am faking the phase shift with the filters on the input of the preamplifier to get this thing to work right. You may see that the phasing does not quite work out on the scope, because you are not actually looking at what I am saying you are looking at, but do not worry about that.)

I am driving the coil now, but at a very low gain on the preamplifier. I am now going to turn up the gain of the preamp—whoa, there it goes. Let's turn that baby down.

I have oriented the phasing so that the force I am putting in is canceling the normal damping in the tuning fork and I get a self-maintained oscillator that generates a loud tone. If I futz with this a little bit—actually, it is already about futzed—I can get this, with the right gain, so that the loudness of the tone is neither decaying nor growing. I have it set just right so that the actively applied "negative viscosity" pretty much exactly cancels the natural passive positive damping.

I can take out a little bit of energy by lightly touching the tines with my fingers, and the sound level drops and stays low. I can put in some energy back in by whacking a tine with my knuckle; I can charge it up more. I can take little bits in or I can take it out, again.

If I turn up the gain the tone starts growing until a nonlinearity limits it.

This is what the prime movers do. Instead of all this electronics, the prime-mover thermoacoustic engines use a temperature gradient and the gas moving in and out of the hot and cold areas to generate this negative acoustic resistor.

If I flip the sign by flipping the input amplifier, notice that the tone decays quite a bit—not really dramatically, because we had it set up so that the active negative damping was just basically canceling the natural damping, so I put in double the dissipation just then and it decayed about twice as fast.

Now I turn up the gain really good. First, let me disconnect the amplifier and you hear it, I will whack it without the amplifier hooked up—that is the normal decay. Now I am going to hook up the amplifier—[taps the tuning fork]—and it goes "thunk." When a thermoacoustic device is running like a refrigerator, the stack or regenerator are pulling energy out of the acoustic wave and turning it into good stuff, into cold six-packs.

This is a good time for a break.

DR. KEOLIAN: You guys made it through the first part of the talk. That was, I hope, the worst part. Some of that was aimed at people already familiar with some of this stuff. The next bit I am going to aim directly, I hope, to people who are not familiar with thermoacoustics and I hope you guys take something away.

During the break, Pat Arnott mentioned that a nice way of looking at those f's, if you just set them equal to zero, is that you get acoustics as you first learned it, no dissipation, just very, very ideal first-chapters-of-Kinsler-and-Frye kinds of acoustics.

If you set the temperature gradient equal to zero, you get ordinary acoustics, but with dissipation. What is new here, what we are talking about now, are funny geometries, but also setting that temperature gradient to something non-zero and you get an active resistance that can be negative or positive resistance.

I also want to thank Tom Gabrielson. A lot of the stuff I said in the first part was thanks to the observations of Tom.

[Transparency 9]

Now we are going to talk about what was originally called "intrinsically irreversible thermoacoustics" by John Wheatley. It is based on standing waves. This was the type of device invented first and it is what many people are still researching, although more may migrate over to the new Stirling traveling wave devices, which we will talk about in the third part of the talk.

The phrase "intrinsically irreversible" comes from using the dissipation inherent in the heat transfer between gas and wall to generate the phase shifts you need to get this thing to work right, whereas the Stirling cycle does not rely on phase shifts generated through intrinsic irreversibilities; instead, the acoustic resonator is made a bit more clever, so it tends to have higher efficiency than the standing wave devices.

I do not think the standing wave devices are going to die off, myself. They tend to be simpler, and simple is good.

[Transparency 10]

Here is your classic thermoacoustic refrigerator—here I am stealing viewgraphs from Prof. Steven Garrett. We have a gas-filled tube, and a driver if we are driving the device electrically.

We will put a stack of plates in there—called the "stack"—and a hot heat exchanger and a cold heat exchanger on either side of the stack.

[Transparency 11]

The stack can be made a lot of different ways. We need a solid material with gaps in it that are about two or threes thermal penetration depth wide. The sound will go through the pores that are a few thermal penetration depths across. In the early days at Los Alamos, where modern thermoacoustics was started, the first thing they did was they stack up flat plates, so that is why we call these things stacks, but you can use other methods.

Prof. Tom Hofler cooked up the method shown here in the transparency of making stacks. It is a long thin strip of plastic with fishing line glued across it. You roll the plastic up into a roll, and you effectively get a bunch of plates that are a distance apart given by whatever thickness you pick for the fishing line.

Another type of stack is this stuff that I'm laying on the projector that you can get from Corning. It is ceramic that has a bunch of little holes in it. This is what they use in catalytic converters in cars. It is cheap, a hundred bucks for a whole bunch of it, more than you could possibly want, and it is much easier than rolling up the plastic.

Also, we heard yesterday that wires work pretty good. You take a bunch of filaments and stick them along the acoustic axis, or stick them across the acoustic axis, or even use an amorphous tangle of filaments. It really does not matter. The point is to have a solid in partial thermal contact with the gas.

[Transparency 12]

The heat exchangers that are on either side of that stack can look like this. This is a stack and heat exchanger from the refrigerator that Steven Garrett and his group over at the Naval Postgraduate School made not far from here. Again, the heat exchangers can be pretty much of any of the classic heat exchangers that you can find. They look like car radiators. This one is a tube-and-fin type; there is a copper tube there with a bunch of copper fins.

[return to Transparency 10]

Let's go back to transparency 10 to see how the thing works. This is the point I want you to take away if you have not seen this stuff before, so I am going to say it about three or four different times until you get it straight, each time a little more and more accurate.

Here is the easiest way to see it—this is what I tell my grandmother. This sponge in my had represents a fluid element. In the middle of the tube, the standing wave causes the fluid element to oscillate back and forth, so I move this sponge back and forth near the center of the tube. On either end of the tube the standing wave compresses and expands the fluid, as does the sponge.

But in between the middle and the end, where the stack and heat exchangers are, the fluid element is doing both. At the stack, when it moves to the left it compresses, to the right expands. When it compresses, it gets hot. When it expands, it gets cold. You put what you want to get cold on the right of the stack. Like a sponge pulling water out of something that is wet, the fluid element pulls heat out of your load. Then the element raises the temperature of the thermal energy through compression, allowing heat to be pushed out at the higher temperature on the left—like squishing the sponge and wringing out the water. That is how it works. That is the first explanation. If you remember that, you are in business. The rest are details.

Why does a gas get hot when you squeeze it? Here is what is surprising. If you flick a ping-pong ball between a wall and a paddle—an ideal paddle and an ideal ping-pong ball—the ping-pong ball will go back and forth. If you move the paddle in pretty quick, the ping-pong ball is going to go faster, it is going to pick up a little speed every bounce because of this moving paddle.

What if you move the paddle really, really slow? You might think if you went really, really slow, you did not pick up much speed, but you have more bounces to go the same distance, which compensates for that in such a way that the speed of the ideal ping-pong ball depends only on the position of the paddle. It picked up the speed of this infinitely slow paddle through an infinite number of collisions, so it speeds up when you squeezed it and slows down as you expand it.

So the little ping-pong ball is going faster, brrrrr, and it is going bang, bang, bang over here. The same thing is happening in a gas being compressed by a piston—or other gas. That is why these things work; you squeeze the gas, it gets hot. Fast molecules are hot molecules. That is the other essential piece of physics.

[Transparency 10]

Now, I lied a little bit. The heat pumping is actually happening near side walls, and the stack is made up of many such walls. Let's say I have a horizontal wall here about where my belly button is and I am going to put my sponge fluid element far away from the wall, up here above my head, so far that there is no time for heat to go between the fluid element and the wall. It is going to be hot on the left, so I imagine there is going to be a hard end of a resonator on the left side (the driver in the picture is at a pressure anti-node, so that end acts as a hard end).

As this fluid goes towards the hard end, it compresses and gets hot, and as it moves away it expands and gets cold. Back and forth—hot, cold, hot cold. It is so far away from the wall at my belly button, though, that this is happening adiabatically, meaning no heat is flowing between the element and the wall.

Now let's move the fluid element really close to the wall, down here by my belly button, and imagine heat (entropy, really) is like water in a sponge, and our goal is to pump water from the right to the left. Let's say that this wall and sponge are wet. My sponge is going to squish out water as it moves toward the hard end on the left. It is in really good thermal contact with the wall, really good wet contact with the wall, so as it comes back to the right, it sucks up just the same amount water from the same place that it just dumped it. It pushes it out and then it just sucks it right up again, as it moves back and forth. This is like isothermal contact—when I have too good a thermal contact between the fluid element and the wall. There is no net water or energy transport.

But now I'm going to move the sponge element up away from the wall a bit, about shoulder high, to get a delay in the movement of water between the sponge and the wall (or heat moving between the fluid element and the wall). I'll get the maximum delayed transfer of heat when the fluid element is away from the wall by about the distance over which heat would want to diffuse in an acoustic cycle, which is this thermal penetration depth.

As the gas about a thermal penetration depth away from a side wall moves toward the pressure anti-node, it compresses, but then it finds itself hotter than the wall. As it lingers towards the left, it will lose heat and compress a little bit more, heat leaving it as if I'm squishing out water from the sponge.

It then moves away from the hard end and expands a little bit and cools. It finds itself cooler than the wall, because those molecules slow down, and it will now suck up heat from the wall. As it lingers on the right picking up heat, the fluid element expands some more. So the fluid element or this sponge goes to the left, squishes out some water, comes to the right, sucks up water, back and forth moving water (or energy and entropy) from right to left.

This would work even if I made it a little bit cooler on the right and a little warmer on the left—or a little dryer on the right and wetter on the left. If I keep doing this, I am going to end up with more water on the left, but I can still keep pumping water up a wetness gradient, up a pressure head, up to a point.

Now it is going to turn out that the temperature swings of each of these little elements might be only about a Kelvin, a degree. We want to pump heat over a greater temperature than this, so let's put many fluid elements in series. Let's have a long row of these moving elements—I now have a sponge in each hand—so they are all going to move to the left and squish out heat, come over to the right, expand, suck up heat, come over to the left, squish out heat, come to the right, suck up heat. I can effectively put a whole bunch of these elements in a series by making a longer and longer stack.

The first element on the right pulls heat from the cold heat exchanger at a really low temperature, brings the energy up to a medium temperature and deposits it on the stack a bit to the left from where it picked it up. The next fluid element picks up that energy from the stack and moves it over to the left some more at a little higher temperature for the next element to pick it up, and so on up to the hot heat exchanger. The function of the stack is to get heat to go from one element to the other; it gives a little temporary spot for the energy to rest so that it can transfer from one sponge to another. It is like a big bucket brigade. That is what many of us call it.

[Transparency 14]

This is described in these fancy viewgraphs that Ray Wakeland made. Blue represents cold, red represents hot. The diagonal lines leaning to the right represent high pressure, the diagonal lines leaning to the left represent low pressure. A big element is low density, a small element is high density. There are three elements going through a cycle in each figure and a wall represented at the bottom. The element furthest from the wall is adiabatic, the closest is isothermal and the middle one is in between. In these drawings, the pressure anti-node (the hard end) is on the right.

First, we are going to have no mean temperature gradient along the stack. When the gas is far enough away to be adiabatic, nothing happens, really. The gas goes back and forth, it is hot on the right side, cold on the left side. For the gas that is really close to the wall, the element goes back and forth but there are no temperature swings. Now there is heat transfer—that is what these big red arrows represent—but the element dumps heat going from left to right and sucks up the same amount of heat going from right to left; so there is no net heat pumping.

But the heat transfer will be delayed a bit for the elements that are about a thermal penetration depth away. The element moves from left to right, compresses and gets hotter. But because of the marginal thermal contact, the heat is being dumped to the wall while the element has reached the right side, rather than on the way there. Next, the element moves to the left, expands, and cools. But it doesn't start sucking up the most heat until it has made it to the left side. The net effect is that heat leaves the left side of the wall and is deposited on the right side. That little delay in the heat transfer is what we need to get the net heat transfer. The delay is coming from the effect of thermal penetration depth. There is a thermal resistance between the gas and the wall, and the element has a heat capacity, so it takes a little time for the element to come to equilibrium as it is pushed about. But we never let it reach equilibrium—we keep it going on its cycle and the heat lags behind.

The heat transfer is intrinsically irreversible, because we are sending heat through a small temperature difference. That creates entropy, which is a bad thing. (The entropy dQ/T_{wall} entering the wall is greater than the entropy dQ/T_{element} leaving the element when $T_{\text{element}} > T_{\text{wall}}$. When $T_{\text{wall}} > T_{\text{element}}$ and heat goes the other way $dQ/T_{\text{wall}} < dQ/T_{\text{element}}$. In both directions entropy was created which ultimately leads to the device's inefficiency.) Much of the inefficiency of the standing wave class of device comes from using this dis-equilibrium of the fluid element to give us the desired phasing for the heat transfer.

We can get a net heat transfer even if we have a small temperature gradient on the stack, as shown in the second figure of Transparency 14. This is the situation we have in a refrigerator—net heat transfer out of something that cold and heat being deposited at a higher temperature.

[Transparency 15]

Now let's put a temperature gradient on the wall or on the stack that just matches the temperature changes of the element as it moves back and forth adiabatically. This is called the critical temperature gradient. Notice this critical gradient does not depend on the amplitude. If we double the acoustic amplitude, the distance the element travels will double, but so will the temperature swings, so the critical temperature gradient is the same.

The critical temperature gradient will depend on where you are in the resonator, however. Near a pressure anti-node, for example, there are sizable pressure and temperature swings, but there is a lot less displacement, so the critical temperature gradient is large there. Closer to the velocity anti-node there is a lot of motion but little pressure or temperature swings, so the critical gradient is small there.

So if we have a temperature gradient on the side wall equal to the critical temperature gradient, nothing much will happen for any of the elements. No matter if they are close or far from the wall, the fluid elements go through the same temperature variations that are already at the wall, so there is no transfer of heat.

What happens if make the temperature gradient even bigger, if we make this right-hand side really hot? For the element that is far away, nothing much happens, as usual. There is no heat transfer because it is too far away. The close by element is isothermal and as boring as it was before. But at the right distance, about a thermal penetration depth away, our little fluid element will come over to the right and compress toward the high-pressure side, but it will find itself in an area that is even hotter than it wants to be. It will suck up heat and expand when the pressure is high, doing work as it pushes itself away from the hard end. It expanded because it got hot—it goes to this hot plate and goes "Whoa, gotta' get away." (laughter)

It then flies away from the hard end, harder than it would have gone had there not been a temperature gradient. It got this extra little kick. It moves away and cools off but it finds itself in a region that is even cooler, so heat leaves it and it cools down even more and shrinks. That creates a small vacuum, which sucks it back toward the hard end, harder than it would have gone.

If you imagine a little fluctuation just starting off in here, a little bit of sound, just a little kick from a molecule maybe, that little fluctuation normally will want to relax back to equilibrium—but now it cannot. When it goes over to the right, it gets hot, it gets pushed away to the left. When it comes over to the left, it gets cold and gets sucked back to the right, and back and forth, back and forth, it keeps growing and growing and grows exponentially as that tuning fork did. It is an active negative resistor. So with enough temperature gradient we get a prime mover that generates sound. The critical temperature gradient marks the boundary between the prime mover and the refrigerator regimes.

Please ask me questions.

MS. HIGHTOWER: Does that all happen in the region of the stack or does this happen along the whole length of the tube?

DR. KEOLIAN: It depends on where the mean temperature gradient is. The sound generation happens mostly in the stack. That is where we put the temperature gradient. It could happen along the walls of the resonator too if there was a big enough temperature gradient there. Let's go back to this picture.

[return to Transparency 10]

Let's say we make the left side really hot, and we make the right side really cold. Now we do not need the driver, we do not need the electricity. We can generate sound with the temperature gradient all by itself. If the left side is hot and the right side is cold, certainly lots of good stuff is happening in the stack.

If the resonator walls are such, though, that they are even hotter to the left of the stack and the tube is even cooler to the right of the stack, then it is also happening along the resonator walls, but there is a whole lot less wall area there, so it is not as important. We can bring the tube to the left of the stack back down to room temperature, if we want, or it can stay hot; it does not really matter too much to the performance of the engine.

That reminds me of something I should have said earlier, which is that if we are making a refrigerator we can get more and more temperature by making the stack longer and longer. But we can get more and more heat power in a refrigerator or more and more sound power in a prime mover by making more and more stacking area. The area of the stack makes the device more powerful. We can adjust the operating temperatures of the refrigerator or prime mover by messing around with position and the length of the stack.

MR. GLADDEN: It seems like the stack spacing is a pretty critical thing.

DR. KEOLIAN: Yes, it is, a very critical thing.

MR. GLADDEN: Is it a couple of penetration depths?

DR. KEOLIAN: Yes. There is a computer code that the Los Alamos guys wrote and they even wrote a manual for it and they give it away for free. What the people in the business do is they use that code and they futz around with the spacing until they get their best performance.

When we went through those equations at the beginning, stepping it from one end to the other, that is what that program does. It ends usually up that the best spacing is about three penetration depths.

MR. TUTTLE: In the middle picture of your standing-wave diagram you have shown a delay. Would you say a little more about the delay?

DR. KEOLIAN: The delay is the critical thing. Remember, there was no delay when the sponge was really close to the wall. When the little sponge squishes out the water and sucks it right back up, again, it does not work. If it is too far away, it does not work. But if I get that delay from being at the right distance the refrigerator or prime mover will work fine. A way of thinking about the delay is that the element has a certain heat capacity, and there is a bit of thermal resistance between that element and the wall. The delay is analogous to the "RC" time of a resistor-capacitor electrical circuit.

It takes time, if I am far enough away, for the heat to get out of the gas and into the wall. In the parts that are really close, the heat goes really fast. In the parts that are really far, the heat takes forever to move. But if they are the right distance away, it takes about a period for the heat to go, or a fraction of period, from the gas to the wall.

That breaks that symmetry between time reversal, you see. If it were just acoustic, going back and forth, you might think that whatever happens ought to be the same way in both halves of the cycle. But there is this net motion of heat from one end to the other. It is from this breaking of the time symmetry caused by that irreversibility that gave the delay, so that when the sponge went to the right and compressed, the heat transferred when the element was on the right, then it came over to the other side and sucked up that heat while it was over on the other side.

MR. TUTTLE: So the thermal transfer holds up the whole thing at one end?

DR. KEOLIAN: Yes, the time it takes for the heat to go from element to wall or wall to element is critical to the operation of this thing. It gives the phasing, the proper phasing that gives us the net heat transfer from one end to the other. If it were not for that, it would not work.

In the other engine that will I show you, the Stirling traveling wave device, we are not going to use this intrinsic irreversibility. We are instead going to use some acoustic trickery to give us the phasing we want.

DR. COSTLEY: So the increase in entropy is bad because it reduces efficiency but, if we did not have it, it would not work?

DR. KEOLIAN: That is correct, and that is why this class of device was originally called the intrinsically irreversible heat engine or, as John Wheatley at Los Alamos liked to call it, the "natural engine." There were no cams or pistons and weird mechanical things like you have in your automobile to get the phasing you need. That is its strength. It is really simple. On the other hand, you are accepting this inefficiency to make the thing work.

DR. ATCHLEY: You are showing the cycle as a series of discrete steps. Isn't the motion sinusoidal?

DR. KEOLIAN: Yes, I do not want to move my sponge like this, squish, move out like that. I want to go Mm-mm-mm-mm, like that. More sinusoidal. (laughter)

MS. PETCULESCU: When you were going from the critical gradient, increasing more and more the gradient, it looks like there is a compromise you have to make, because if you increase more and more the gradient, then you will do worse and worse in generating entropy.

DR. KEOLIAN: Correct, but I will get more and more heat.

MS. PETCULESCU: Yes, so you have to stay somewhere where if you want reasonable efficiency, you do not have to....very much.

DR. KEOLIAN: Gabriela is absolutely right. There is a tradeoff between efficiency and power density. We get the best efficiency when we are at the critical gradient, but we get no heat transfer, we get no sound generated. This is generally true with all heat engines, it seems.

But if we go off equilibrium, if we go away from that critical gradient, we can get some heat going but we are going to pay for it with inefficiency. So there is a tradeoff between making a really efficient engine and making one that is compact. If I can get more heat from a given stack, I do not have to make such a big heavy engine.

There is no best engine, there is no best design. The user gets various tradeoffs between whether she wants high efficiency or he wants a really compact one. The user decides which one is the best.

DR. HAMILTON: What determines where you put the stack, because I would think there is some gradient there, and your biggest gradients would be at a quarter of the tubes, it always seems to be toward the rigid end.

DR. KEOLIAN: Yes, Mark brings up a good point, which is we need both the pressure swings to get this sponge to expand and contract and we need the motion; it is quadratic in the amplitude. Where in the resonator do we get the most of the product of those two things?

[Transparency 10]

At the hard end of the tube we get all the pressure swings—and thus temperature swings—but no velocity swings. In the middle of the tube we get all the velocity swing but no pressure swing. We need both. The biggest product of the acoustic pressure and velocity is at a quarter of the way along the tube, or an eighth of a wavelength from the end. But if you look at these engines, we never have the stack there.

The reason is that this gas moving back and forth is rubbing against the stack walls. The gas is viscous, and that is giving us loss. So we generally trade some of that velocity for more pressure—we trade some of the power density for better efficiency—and push the stack closer to the hard end than a quarter of the tube to get the velocity down. We get a little bit more in the temperature swing, but not much, but we gain on the efficiency.

The computer program that came out of Los Alamos, DeltaE, that is used by a lot of people—you just futz around until you get the best compromise.

MS. PETCULESCU: And the lower Prandtl number should not matter very much?

DR. KEOLIAN: No, it does. Lower Prandtl number is a good thing. Lower Prandtl number means we get that heat transfer between the sponge and substrate, but we do not pay so much in the viscosity of the gas moving back and forth, so we might be tempted to bring the stack a little farther away from the wall.

MS. PETCULESCU: [Inaudible]

DR. KEOLIAN: What Gabriela is saying is I previously showed that the magnitude of the good thermoacoustic resistance did not seem to depend too much on the Prandtl number, so she is wondering why would I say that we might want to move the stack away from the hard end if the Prandtl number were improved, which we can do, say, by adding some argon or xenon to the helium.

The viscous losses still depend on the viscosity and the velocity, so if my viscosity is less, we are able to accept more velocity and get the power density up, get more motion out of the gas, so I think we am still tempted to move the stack closer to the velocity anti-node.

It turns out there are so many variables in this business, it is really hard to wrap your brain around them all and keep track of them. So if you are getting confused, do not worry about it, everyone does. That is why the DeltaE program and others like it were written; you just try a bunch of things, to some extent. If you have been in it for a few years, you can sort of do this in your head, but it is really hard.

DR. MIGLIORI: In other engines that have intrinsic irreversibilities, for example, thermoelectric coolers, in which heat conduction is always present, there is a critical temperature gradient in which you put in work and the engine pumps heat from hot to cold, so it does nothing. There is a sort of dead band between where it is acting as a refrigerator and where it is acting as an engine.

DR. KEOLIAN: Thanks. The same is true here too, but I've been ignoring it.

[Transparency 16]

If we can get a temperature gradient bigger than that critical gradient, this thermoacoustic device generates sound. If it is less than that, it generates cooling. Look at this cool thing we can do. This was called the beer cooler at Los Alamos long, long ago. It is a Hofler-style refrigerator; Tom Hofler was a grad student of John Wheatley's (he is over the Naval Postgraduate School now). He made these quarter-wave kinds of pipes with this big sphere on the end.

The main thing I want to show you is that you can combine a prime mover and a refrigerator to make a heat-driven refrigerator. You can make the hot heat exchanger end of this device really hot, with a flame, say. You make the middle near ambient temperature, cooling it with ambient water or ambient air. This big temperature gradient in the prime mover stack generates sound.

You can then use that sound to pump heat from the cold exchanger up to the middle, room temperature exchangers. This second stack on the bottom here is a refrigeration stack. The upper one is an engine or prime-mover stack. The net result is that heat comes in to the hot side and comes in from the cold side and leaves from the middle. Both the flames and the bologna are getting colder and the atmosphere is getting hotter.

[Transparency 17]

One of the things we are trying to push at Penn State is a geometry like this, here's a cartoon of it. One of the reasons I'm showing you this is to let you know that you can use a radial geometry for thermoacoustic devices as well. Let's say we have a duct of hot exhaust from something, like an automobile exhaust, diesel exhaust, or gas turbine exhaust. Let us wrap the duct with this contraption.

We bring in some of the hot exhaust and get one side of this innermost stack hot. We can then cool this second-most outer place to ambient temperature with some ambient fluid, put a stack in there, and generate sound. We then get a radial sound wave being generated. Gas comes in, gets hot, jumps away, cools off, comes back in, generating a large amplitude radial mode.

Toward the outside of this thing is a refrigerator. As the gas moves to the outside it gets warm, dumps heat in the outermost ambient heat exchanger, comes back, expands, sucks up heat from the cold exchanger,. So we have a heat-driven refrigerator. We cool the exhaust, cool the six-packs, and something else gets warm, like the ocean or the atmosphere.

Now let's get a computer involved. I'll be showing you many of the computer animations written by Greg Swift. You can download these animations from the Los Alamos thermoacoustics web site http://www.lanl.gov/thermoacoustics/. [Computer demonstration]

We will look at the animation called WAVE first. This is a Macintosh, running Windows, running DOS. (laughter)

What this is showing is a wave in a tube. We have pressure on the top, velocity on the bottom, and little vertical lines that represent the boundaries of fluid elements—you can think of lines of paint on the fluid elements—and these things are moving back and forth.

You can see a traveling wave here. It is a wave traveling to the right. On the bottom of the animation is the equation describing the pressure. It is a constant plus $R \cos(\omega t - kx)$, which is a wave going to the right, plus $(1-R) \cos(\omega t + kx)$, which is a wave going to the left—and I can vary R. Right now R is 1, so it is all traveling wave going to the right.

I can type a v and a little dot shows on a couple of places on the waving line. The vertical position of the dot is proportional to the pressure. The horizontal position of that dot tracks the vertical lines that map out the boundary of a fluid element. The phasing is such that the dot is sweeping out an ellipse.

If you like, you can think of each of these fluid element boundaries as being like a piston. Consider this boundary below this right-hand ellipse. Think of all the stuff to the right as being like a volume, the boundary as being like a piston, and that this fluid element is compressing the volume of the stuff to the right.

The ellipse then is mapping out the pressure p and volume change dV of the stuff to its right. The area inside the ellipse is the integral of $p\, dV$, and that is the net work that is being done by our fluid element on all the fluid to its right..

Mapping out the ellipse in this way shows that the pressure is low when it is moving to the left, and the pressure is going high when it is moving to the right. If you think of this as a piston, it is doing more work when it is moving to the right, because the pressure is high when it is moving to the right, than it is getting back when it is moving to the left. So there is net energy going from left to right. You can see that from the area of that ellipse. When it is being traced out going clockwise, that means energy is going from left to right.

If I change that R factor to be 0.5, which means I have an equal amount of wave going to the right and an equal amount going to the left, I get the standing wave. Now take a look at this little pV indicator diagram, as the ellipse is called. It is a straight line now, there is no area in it.

Let's look at the left indicator diagram first. For the one on the left, when the element is moving to the right, the pressure goes up. That is what would happen if we were pushing against a spring. When we move back, we get all the force right back, again, all the energy we put into the spring comes right back out again. This is a good spring when we have a very good standing wave.

But notice that with the indicator diagram on the right the pressure goes down when element moves to the right. The gas to the right of the element is mimicking a mass. The standing wave on the right-hand side of this velocity node looks like a mass. On the left-hand side it looks like a spring. It is purely reactive either way—some nice acoustics there.

If I set R equal to zero, we have the traveling wave going to the left and now the ellipses move in the counterclockwise direction, showing that energy is moving from right to left. If I make R equal to 0.6 now, a little bit of standing wave and a little bit of traveling wave, I have an imbalance between the right- and leftward-going waves.

What I want to show you is that these ellipses become sort of tilted. I am going to need this to explain the other things later in the day. The area of it tells you how much traveling wave there is. The slant of it tells you about the standing wave piece of it. If it is moving clockwise, energy is going from left to right. If it is going counterclockwise, energy is going from right to left.

MR. APOSTOLOU: What are you going after with the work?

DR. KEOLIAN: Right now I am just talking about some basic stuff. I have no goal at this point other than to set you up for the next animations. When we put a thermoacoustic device into the animation we'll want to see how the work flows.

We have been talking a lot about penetration depths. Now I am going to show it to you, finally. First, we are going to do this oscillating wall animation, OSCWALL. If I have a fluid at rest at infinity and a wall moving back and forth, the fluid (meaning either a liquid or a gas) right next to the wall is stuck to the wall by viscosity and it has got to go back and forth with the wall.

Far away from the wall the fluid is not moving, but there is a distance over which there is some influence of the wall. Think about the physics of it. If the viscosity is really big, then the wall has an influence that reaches out farther into the gas. If the inertia of the gas is really big, then it limits how far the influence can go, because there is a balance between the viscosity of the gas trying to get the fluid away from the wall to move and the inertia of it trying to keep it from moving.

There is a characteristic distance, the viscous penetration depth, which depends on the ratio of viscosity to density. It also depends on frequency. You can see a wave traveling away from the wall out to infinity, but it is highly damped. It has a wavelength—Prof. Atchley said this yesterday nicely: The real part and the imaginary part of the wave vector are equal.

This characteristic distance, each of the tic marks on the vertical axis, is a viscous penetration depth. When you get to about four penetration depths away from the wall there is not much happening at all.

Sound behaves like this, except in sound the fluid is moving and the wall is stationary. That is shown on the next animation, VISCOUS.

Here we have a piston on the right-hand side of a tube driving the gas. The piston is not moving very much. Hitting a key has the animation plot the velocity and the pressure. This is sort of like basic acoustics now. It appears that the velocity goes to zero at both ends, but it does not quite. If you look carefully, there is a little bit of velocity that has to match the piston velocity. The velocity swings in the middle are much bigger than the velocity of the piston. That is the "Q amplification," because we are near resonance.

Next, we turn on the vertical lines and we can see the gas in the middle sloshing back and forth—a typical standing wave. But hey, these little lines are rubbing against the wall. We are showing them moving, but that cannot happen, because right next to the wall they're stuck by the viscosity.

So let's blow up a region in the center of the tube right near the wall and zoom in on it, and see what is happening. Not too far from the wall, a few viscous penetration depths, the gas is moving like a big plug of fluid, like we just saw in the larger scale picture. But right near the wall there is a viscous wave being launched by the wall going into the gas, like we saw in the animation VISCOUS. You can think of it as the viscous wave canceling the acoustic wave right at the wall. It is adjusted in just the right way such that those two solutions add up to zero at the wall and you get this neat little motion from the combination. Notice it is coming to the right and then sort of whips along like that, and that's what's going on within the viscous penetration depth.

MS. HIGHTOWER: Why does it....near the wall but not right at the wall? It looks like it goes faster than the liquid.

DR. KEOLIAN: Yes, isn't that neat? It seems to lead, doesn't it?

MS. HIGHTOWER: Why?

DR. KEOLIAN: I do not know why, but it is neat. (laughter)

There has to be a physical explanation, but I do not have it. That is what students are for. You have the time to ponder those things.

Now let's go back to the oscillating wall demo, OSCWALL, which I already showed you. But now let's imagine that instead of the wall oscillating, the wall is fixed, and imagine that what we are plotting here is the temperature of the wall and the gas. We have a wall whose temperature is oscillating up and down. Right near the wall the gas has to follow the temperature swings of the wall. Far away the gas comes to whatever temperature it is at infinity, which is some constant. We can see that there is a thermal wave that goes from the wall out into the gas.

This is like what happens to your water pipes. You can think of it as being hot when the base of the wave is to the right, as in summer. Then it goes to fall in the middle, and then winter when it goes to the left, and then spring and then summer. There is a thermal wave in the soil that goes down, away from the surface (up in the animation). For every place in the country you have to bury your water pipes down to a certain depth so that they do not freeze, and when they freeze is not in the winter.

Notice that if you are some depth away, there is a delay. Like right now in the animation it is winter, and two penetration depths away, say, the buried pipe is still cooling. A moment later it is spring and, bingo, two penetration depths away, the pipe is coldest in the spring, not in the winter.

The same thing happens in acoustics. Let's go now to the animation THERMAL. This is similar to what we had seen before. We have a standing wave in a pipe and pressure swings that are biggest on the left-hand side of the wall. We are also going to plot the temperature now. The upper graph is the temperature and the temperature swings pretty much follow the pressure swings.

We are going to look at the area right near the wall. We are showing the temperature swinging right near the wall, but that cannot be true, because the wall has a whole lot more heat capacity than the gas does. Actually, what is important is the amount of heat capacity within a thermal penetration depth in the wall—the wall has its own thermal penetration depth, too. It is the heat capacity in that little accessible volume at the end of the wall compared to the heat capacity of the gas within its thermal penetration depth. Since the wall is a thousand times denser than the gas, it has a lot more heat capacity and so remains isothermal.

So let's zoom into the area near the wall. Right near the wall the temperature cannot swing very much and that is shown in the animation. You can think of that thermal wave being launched from the wall into the gas in such a way that it cancels the temperature swings that the gas was imposing on the wall. The diffusive wave of heat is going from left to right, and that is what it looks like in the animation.

This leads to dissipation of acoustic power but that is a little hard to see. This is the cause of that thermally induced acoustic resistor r_k in that equivalent circuit we had way back when in the first hour. We can see that over here on this bottom graph. Here we have the pressure and the fluid elements' boundaries shown, so we are going to consider one of the fluid elements as a piston and ask how much net work it does it transfer to the remaining fluid to its left.

We are going to plot out the pressure and position with one of these little ellipses as we have done before. It is basically a straight line as I showed you for the standing wave. The gas to the left basically acts as a spring. Look at it from the right to the left. Imagine you are standing on the right of this little piston and you are pushing in. The pressure is going higher as you are compressing the volume on the left, and the pressure is high again as the piston comes back, again, but the pressure and motion are not quite in phase. The pressure is a little bit higher on the way in than on the way out. Why is that?

There is a little bit of area to this ellipse, which means some work, real live time-averaged work, is going from right to left. The reason is, when the gas moves in, most of it is trying to compress adiabatically, but in this region—a thermal penetration depth away from the wall—the gas cannot make up its mind if it is adiabatic or isothermal. On the way in, the gas tries to compress adiabatically and raise its temperature, which causes a little bit of heat to be transferred to the wall once it's compressed. On the way out, it tries to adiabatically expand but it is now a little cooler at any position than it was on the way in because it had just lost some heat. After it expands it pulls some heat out of the wall and is back where it started.

The gas basically acts like a gas spring and the effectiveness of that spring depends on the temperature of that gas. When the molecules are moving faster, they push against the piston harder. It is a stiffer spring when the gas is hot, so it is a little stiffer coming in than when it is coming out. The piston put work in when compressing the spring, but it did not get quite all that work back out again as the spring expands, because the gas was a better spring coming in than when it was coming out. Where did that work go?

It went into the wall as a net heat, as a net flow of heat from the gas into the wall, in addition to the oscillatory heat flow from the temperature oscillations of the gas. Entropy is being produced by the oscillatory heat flow because heat is flowing through a temperature gradient, giving an irreversibility that leads to a net loss of work and a net flow of heat out of the gas.

The net work that is passing from right to left is given by this third line and the derivative of it tells you where the most dissipation was occurring. It is occurring about a thermal penetration depth away. That is where the net work is being lost from the acoustic wave and being turned into a net heat.

You guys are nodding off. I am going to have to give you a break here, I think, and we probably should open some doors.

DR. KEOLIAN: I want to finish up on these intrinsically irreversible standing-wave engines and then move on to the Stirling engines. One last standing wave demo, called STANDING.

[Computer demonstration]

Here is pretty much what we had before, a standing wave in a tube with velocity and pressure—shown on the two moving graphs. Now I am going to put in a stack of plates from the hard end and we are going to zoom in on that little bit in-between the two plates and watch a little blob of fluid going back and forth.

PARTICIPANT: On this one it says "parcel."

DR. KEOLIAN: Parcel on this one. Sometimes it is a blob, sometimes it is a parcel, sometimes it is a fluid element.

What we are watching is the pressure and volume of the parcel. Now notice, that this volume is the volume of the parcel, not the volume of everything to the left. It is a straight line right now, because I did not push the right button. (laughter)

Let me start over here, and type an *f* for a flat temperature profile on the stack. What we are plotting is the temperature of the blob of gas when there is a flat temperature profile, no gradient, on the stack. The little blob is moving back and forth. When it gets hotter than the stack, heat leaves the blob and goes into the stack, as shown by the red arrows. When it is cooler than the stack, heat goes from the stack into the blob.

A consequence of that is that the pressure is higher when the gas is warmer and compressing and a little lower when the gas is expanding and coming out, so there is a net area to this pV diagram for the blob. The little blob is absorbing work, which is a consequence of pumping heat from right to left.

If you think back to those first thermodynamic diagrams I showed you, with work, cold heat and hot heat, you cannot pump heat without doing some work, although right now we are not pumping over a temperature gradient, so the Carnot efficiency would be infinite.

Let's now put on a temperature gradient by typing *r* and have the blob pump heat against some temperature gradient. This is now a refrigerator—we now have a slight temperature gradient. The slope of this ellipse does not change. That is determined by the acoustics, by the ratio of pressure and velocity at that point.

But we still arranged it so that the blob is a little bit cooler than the stack when it moves to the right and a little bit warmer than the stack when it moves to the left, so to the right it absorbs heat from the cold end, and dumps heat towards the hot end, and that is a very good thing. That makes it a refrigerator and refrigeration is a good thing to have. It is actually a $750-billion-a-year industry worldwide, that is how good a thing it is.

MR. PORTER: The area....got less. Is one more efficient, like there is a difference in efficiency based on what that area is?

DR. KEOLIAN: Yes. Part of it is efficiency, part of it is Carnot efficiency.

Let me go to the next one, the critical temperature gradient by typing a *c*, and what we have done is adjusted the temperature gradient so it matches the temperature swings of the particle. Notice, though, that the area went to zero and that is how I know that in this simulation Greg did not throw in the viscosity of the gas.

MR. PORTER: So right now it is just basically a thermal gradient, not a viscous gradient established with the boundary layer?

DR. KEOLIAN: I hope I am not saying this badly, I do not want to get you guys confused—yes. Ignore what I just said. (laughter)

The area goes down because I am not pumping any heat and either the viscosity is neglected or for some reason it does not affect it here. The viscosity of the gas is not making the ellipse open up like I think it should.

DR. COSTLEY: In that case, if you had viscosity you would be putting work in.

DR. KEOLIAN: That is what I am not a hundred percent sure of. I think I have to put work into the blob because the blob is going to be doing work against the wall. No, it does not do any work against the wall, excuse me, because the wall does not move. It takes work, turns it into heat and heat goes from the blob to the wall in a net way. So I'm pretty sure that this simulation is in the inviscid limit—we are imagining the Prandtl number is zero.

The critical temperature gradient is the crossover point between a refrigerator and a prime-mover engine, which I will show you next by typing a *p*. Now we put a hot temperature on the left of the stack. The blob picks up heat when the pressure is high. It dumps heat when the pressure is low. It expands pushes out from the end of the pipe when the pressure is high, so it does work. It comes back relaxed, so the blob is doing net work. The net work is showing up on the pV diagram and that work is turning into sound. That is good stuff— the engine is generating sound.

[Transparency 19, 20]

Just to show you that this really happens, we have a real live demo here. What I have here is what I believe to be the world's simplest thermoacoustic demo. It was made by some students over at Penn State, primarily Reh-Lin Chen, and also Kevin Bastyr had something to do with this one. This one is Kevin's version.

The Corning ceramic material is used as a stack in a test tube quarter wave resonator. A little nickel-chromium wire is wound on the end of the stack facing the closed end of the test tube. The nickel-chromium wire is acting as the hot heat exchanger and is attached to copper wires which come out of the open end of the tube. Reh-Lin has also driven this demo by focusing sunlight on the hot end. You do not need any electricity.

There is, in effect, not only no hot heat exchanger, there is no cold heat exchanger. What would be the cold end stays cold, perhaps because there is convection in here or maybe it is just the heat capacity of the thing.

I am going to power up the nickel-chromium wire with a couple of six volt lantern batteries, and lots of sound comes out. It is not real touchy as to where I put the stack, which I can slide around in the test tube with the copper wires (but it is touchy about that wire being broken). If I get it too close to the hard end, it does not work, or too close to the open, it does not work. It has got to be just right.

That is it for the standing-wave engines.

[Transparency 21]

Now we are going to turn our attention to the recent development, the Stirling class of engines. They are kind of traveling-wave-like and you will see what I mean by that. Let's do that with these animations. We are going to go to the one called PTR, or pulse-tube refrigerator. [Computer demonstration] Some of the names make no sense at all—"pulse tube refrigerator" is historical.

A Stirling engine is quite different from a standing wave engine. This is a classic old refrigeration or heat engine cycle. What we have are two pistons moving back and forth and a stack-like thing is in between, except that it is going to be called a regenerator now . The difference between a regenerator and a stack is that the regenerator has really small little pores, much smaller than a thermal penetration depth.

The typical way a regenerator is made is by getting a lot of window screen, except really, really fine window screen, cutting it up into a bunch of circles and stacking those circles up against each other. It makes a real tortuous little path for the gas to go through, getting the gas in really good thermal contact with the secondary thermodynamic medium.

I am going to slow this animation way down so we can talk about it as it is happening. On the top is plotted pressure. On the bottom is temperature. What is happening now is that the two pistons are moving to the right, so we are pushing the gas from a warm temperature, room temperature, down to a cold temperature.

Then we let the gas expand. We move the pistons away from each other, and everywhere between the pistons the gas expands and either lowers its temperature if there is nothing nearby, or it sucks heat from whatever there is around it. In particular, the gas to the right of the cold exchanger has expanded and turned cold.

The next thing we do is we start pushing the gas to the left by moving both pistons to the left. The cold gas on the far right comes in and smacks into the cold heat exchanger on the right and cools it down, pulls heat out of it. Meanwhile, all the gas in the regenerator is moving to a warmer area and pulling heat out of where it is going.

Then we start compressing the gas everywhere. The gas starts dumping heat into whatever it can, because it is trying to get hotter. The gas to the left of the hot heat exchanger does get hotter, because there is not anything near enough to dump heat into.

Then we start displacing the gas to the right and that hot gas that was over to the left starts dumping heat into the hot heat exchanger.

The net effect is that we cooled the right-hand exchanger and heated up the left-hand exchanger. We pumped heat from a cold temperature to hot. It is a nice little refrigerator. In the animation, we are expanding the gas now. This gas over here to the right of the cold heat exchanger is getting colder. It is lowering its temperature below the temperature of the exchanger. Then we shove all the gas to the left, that cools off that cold exchanger. The next thing that will happen is we will compress the gas, then we will heat up the gas on the left—I will speed it up a little bit—we are displacing to the left.

Let's zoom in on what is happening near that cold exchanger by typing a *c*. Here is a little blob near the cold exchanger. The gas is moving to the right. We then expand it, it gets cooler, smacks into the cold heat exchanger, dumps heat into the exchanger. The blob hits the exchanger, changes its shape to get in there, it is colder, starts pulling heat out of the exchanger—that is a good thing, that is what we are paying for it to do. It then goes into the regenerator, which has a temperature gradient on it.

It is pulling heat out of the regenerator as it is moving to the left, because it is coming from a colder area moving into a hotter area, but then when it comes back, again, it is dumping all that heat right back, again, just like that sponge I showed you that was too close to the stack. It pushes the water out and then sucks it right back up, again.

It goes through the exchanger, pops out like that. It is now kind of floating out there. The pistons are now moving apart, it is expanding, it is dropping its temperature, and then coming back to complete the cycle.

DR. COSTLEY: What part does the regenerator play?

DR. KEOLIAN: I will show you. It is, again, sort of like a bucket brigade. Each parcel is moving back and forth. That is actually a hard question. When the heat pumping is zero, exactly the same amount of heat is moving to the left as is being sucked up on the way back, but that is not exactly true. There is an ever-so-slight imbalance between the two, and the heat is moving from right to left as a bucket brigade all the way down. There is a lot of heat, because there is a lot of heat going back and forth between the gas parcels and the regenerator, because they are in really good contact. *[What I said in these last three sentences is wrong, and caused much confusion in the following discussion. I was confusing entropy flow with heat flow. What I wish I said is this: From what I can tell, the regenerator does two things—it acts like a bucket brigade and it changes the temperature of the gas so that it can accept heat at one temperature and reject it at another. The cycle of a blob of gas that's inside the regenerator can be seen in the animation by typing an* r. *When the gas is towards the right it is made to expand by the two pistons moving apart. This causes the gas to pull heat from the regenerator. Next the pistons move to the left pushing the gas into a warmer area, absorbing some more heat and increasing the blob's temperature. The pistons then move towards each other, compressing the gas which deposits heat on the left. Lastly, the pistons move to the right, the blob goes down the temperature gradient replacing the heat it took out of the regenerator when it was moving to the left, and lowers its temperature, ready to start again. Notice that the regenerator was able to raise and lower the temperature of the gas without requiring work from the pistons (ignoring viscosity)—ideally the gas motion occurs at constant volume so it does not absorb work. Notice also that heat was pulled out when the blob was on the right and deposited when it was on the left. The next blob to the left of this one will do the same, and so on, setting up the bucket brigade that pulls energy out at a low temperature and rejects it at a high temperature.]*

DR. MIGLIORI: ....the gas molecules make it through without transporting heat through this thing so you can get the gas from cold to hot without having to warm it up—

DR. KEOLIAN: But I could do that with a big chunk of Styrofoam. I have to do more. I have to pump heat up the temperature gradient, but I do not want any conduction.

DR. MIGLIORI: But it is a Stirling engine, right? The function of the regenerator is to allow you to get the gas to the hot end and have it be hot when it comes out, and then all the action occurs in those models, compression and expansion.

DR. KEOLIAN: What you say is true, too, but there is still heat-like stuff, there is still a flow of —

DR. MIGLIORI: But ideally there would be no flow of heat through the regenerator.

DR. KEOLIAN: That is not true. There would be no conduction, but there is a flow of heat going one way that is balanced by a flow of work going the other way, so that the net energy is zero, but that is for the cognoscente, this argument. *[Better stated in retrospect: There is a flow of entropy (not heat) in the regenerator. Even though the blob is at a particular temperature determined by where it is in the regenerator, when it is moving to the left it is less dense than when it is moving to the right. When it is less dense it carries more entropy (at the same temperature), so more entropy is moving to the left than is moving to the right.]*

DR. GILBERT: It seems like I remember Greg Swift was saying we want this to be an isothermal process so that there is no entropy *[generation]*, that you do not have anything flowing through a temperature *[gradient]*—

DR. KEOLIAN: Yes, that is true. The gas is in such good thermal contact with the regenerator that if we look at one place in the regenerator, the temperature does not swing there. However, if we follow an element, that element is going through temperature changes. It is the difference between a Lagrangian and Eulerian way of looking at it.

DR. ARNOTT: But it is always looking at something that is the same temperature as it is.

DR. KEOLIAN: Yes, that is also true. As the gas moves back and forth it is changing its temperature. But it is in such good thermal contact that we are not creating much entropy in getting heat to flow back and forth between the blob and the regenerator—an ever so small amount because the temperature difference between the blob and the regenerator is so small. It is not intrinsically irreversible the way the standing wave devices are. There are some irreversibilities if we pump large amounts of heat but, if we do not, then the efficiency approaches Carnot's efficiency.

Let's look at it from a sponge's point of view. I am going to bring a compressed sponge to your right, expand it, then move it to your left and then compress it—expand it on the right, compress it on the left. If we imagine that the sponge is in contact with something wet, it will pull water out of the right and bring it to the left, building up a puddle of water on the left. Because the sponge is bigger as it moves to the left it can carry more water to the left than it brings back coming to the right.

In the standing wave devices, the important blobs were a thermal penetration depth away. The blob moved to the left towards a pressure anti-node, compressed a little bit and raised its temperature as it did so, then squished out a little bit more as heat left it and it thermally relaxed. Then it moved to the right, expanded quasi-adiabatically a bit and cooled, then expanded some more as it pulled heat out of the stack. It expanded on the right and compressed on the left because of the marginal thermal contact. Now with the Stirling cycle maybe you can see that the blob ends up doing essentially the same thing, but instead of relying on thermal relaxation to give us the phasing, we are forcing the expansions and contractions of the gas by moving the pistons apart or towards each other.

The effect is the same, I want to expand over here and compress over there. In one case I do it with the standing wave and its thermal delay. The other way I just brute force it with the pistons.

DR. GILBERT: What is the reason? What are you getting by doing it the Stirling way?

DR. KEOLIAN: We are getting the same effect—we are getting net heat going up a temperature gradient, which is worth $750 billion.

DR. GILBERT: It is more complicated device. There must be some advantages to this in efficiency.

DR. KEOLIAN: We did not have to have the heat flowing through a temperature gradient between the blob and the regenerator. We did not have the intrinsic irreversibility of the standing wave device to get the device to work. We are going to impose the phasing we want, in this case, with levers and pulleys.

DR. GILBERT: Well, that makes that Stirling device more efficient.

DR. KEOLIAN: Correct.

DR. ARNOTT: You are still sloshing the gas through these narrow *[channels]*—

DR. KEOLIAN: Yes, and that is a problem that we will get to—I hope.

Notice that with the left-hand piston, the pressure is high when it comes in and low when it goes out. Work is going from left to right. With the right-hand piston, the same thing is true; work is going from left to right. The right-hand piston is absorbing work from the stuff on the left. The left-hand piston is doing work on the right hand piston.

In the traditional Stirling engine this work going into the right piston is recovered with levers and cams connected to the left piston to help push it; in any case, that energy was not lost.

But there are levers and cams and junk that make it complicated, which is not that much fun. Then somebody figured out that this piston over here on the right is just absorbing work. If we put a big absorber over here, we can do the same thing.

A way of doing that is with something called the orifice pulse-tube. We can show it on the animation by typing an *o*. Here is the orifice. It is basically a big acoustic resistor. There is a big volume behind it that is basically a big spring or compliance. If you like, you can look at it as the analog of an electrical RC (resistor-capacitor) network. The combination just absorbs energy.

The orifice, because it is absorbing energy, is warm, and we are trying to make the cold heat exchanger cold, so we want to keep these two things away from each other. We could do it with a big block of Styrofoam moving back and forth or we can do it with a long enough tube of gas. This tube between the orifice and the cold heat exchanger does the thermal isolation. A good name for it would be a thermal buffer tube but historically it has been called a pulse tube. We are losing energy in the orifice, but we got rid of some parts, which is good.

[Transparency 23]

What an actual orifice pulse-tube refrigerator looks like is something like this. You have a regenerator with a cold exchanger on the top and a hot or ambient exchanger on the bottom, a pulse tube, which is basically a tube at least about three displacements of the gas long (that is about what is needed to keep them isolated), a valve that acts as the orifice so you can adjust the amount of dissipation, and a tank so the gas has somewhere to go after it passes through the orifice.

Los Alamos figured out to make it a little bit better. They can mess with the inertia, as well as the resistance, in this newer version with another valve. So that is really swell.

DR. GARRETT: Do you want to say that the problem is that the....power that is being absorbed is not being fed back?

DR. KEOLIAN: Yes, this power that is being absorbed is just wasted. Nevertheless, it is not a bad thing, because you can now use a conventional thermoacoustic driver to make something useful. [Computer demonstration]

Look what we can make, shown by the animation TADOPTR. It's a Thermoacoustically Driven Orifice Pulse Tube Refrigerator. We can use a stack to generate sound, get its left-hand side really hot, have the right hand side of the stack and the rest of the tube at ambient temperature, put in our regenerator over here to the right, a pulse tube or thermal buffer tube next, then an orifice to dissipate the energy, and finally a volume as somewhere for the gas to go behind the orifice, over there. The left end is really hot with flames, the middle is at room temperature, and the cold exchanger on the regenerator is really cold, in fact, cold enough that it can liquefy natural gas.

What Los Alamos is doing is building a gadget like this that will liquefy natural gas. Often, natural gas is found in oil wells, but usually not enough to make it worthwhile to pipe it out. It is too expensive to make the pipe. So what is often done is the gas is just burned up near the well to get rid of it, putting all that carbon in the air and just wasting all that juicy energy. What the Los Alamos guys are doing is burn some of that gas, make this left end hot, keep the ambient end cold (I think it is being cooled by the gas coming in, but I am not sure about that), and liquefying the rest. They drive the liquefied natural gas away in trucks and make money.

The animation shows it working. It sort of puts together the things we have been talking about so far. The left-hand side looks like the standing wave devices we talked about in the second part of the talk, the right-hand side is a Stirling, traveling-wave device that we have just been describing.

But as Steve pointed out, we are losing energy in the orifice and this had been bugging Swift for a little while. In the meantime, they put out this cool video. [The video was shown.]

MR WOLIAN: "I am John Wolian, director of the acoustic liquefier program at Cryenco here in Denver. We are going to show you a totally new technology for the liquefaction of natural gas. We expect this technology to open up entirely new markets and applications for liquefied natural gas."

DR. SWIFT: "In 1989 a small group of us at Los Alamos and at the National Institute of Standards and Technology in Boulder invented the first cryogenic refrigerator with absolutely no moving parts. Our first experiments....but the hardware had very little refrigeration power and it was not very practical.

"When Cryenco heard about our invention, we began working together to make it powerful and practical for liquefying natural gas."

WOMAN NARRATOR: "Natural gas is an abundant fuel widely used for home heating, for electric power production, as a heat source in industrial applications, and as....for chemical processes. Most natural gas is carried from gas wells to the consumer in pipelines. Where pipelines cannot be used, natural gas is transported and stored as a liquid, because it takes up remarkably less volume.

"Natural gas can only be liquefied by extreme cooling to about minus 140° Fahrenheit, much colder, even, than dry ice. When liquefied, the gas is known as LNG, liquefied natural gas. Such cold temperatures are called cryogenic temperatures. Until this project, all cryogenic liquefiers had moving parts.

"Moving parts are subject to wear and leakage and often require lubricants and they may be costly to manufacture. Our new liquefier avoids these problems. It has no moving parts and may begin a major revolution in the use of liquefied natural gas."

DR. SWIFT: "This is the world's first acoustic liquefier. It has three major components, a natural gas burner, which provides the input power, an engine, which converts the heat of combustion into an oscillating pressure wave inside this pipe. The oscillating pressure wave drives the cryogenic refrigerator, which liquefies natural gas at minus 240° Fahrenheit at a rate of 100 gallons a day or about a cup a minute, so we put energy in the form of heat into that end and we get refrigeration power at this end. All this happens with no moving parts.

"The hardware is simple, but the physics is complicated. At a pressure of 30 atm this helium is the thermodynamic gas for both the engine and the refrigerator. The helium pressure oscillates in a half-wavelength acoustic wave like the sound wave in an organ pipe.

"The pressure goes up and down in the two ends of the pipe as the helium back and forth near the center. On the left, the engine uses the heat from the burner to generate the acoustic wave. The thermal expansion and contraction of the helium pushes and pulls on the acoustic wave, just like thermal expansion of the combustion products in a car engine pushes on the pistons.

"While the engine pumps acoustic power into the wave, the refrigerator takes acoustic power out of the wave. Here, every time the pressure and the acoustic wave goes down, the helium expands and cools. That cold helium liquefies to natural gas. The helium oscillates 40 times a second here.

"All that action in this simple hardware, a pipe full of helium, three heat exchangers on the one end and four heat exchangers on the other end, and no moving parts, no sliding seals, no lubricating oil, nothing to wear out."

MR. WOLIAN: "Cryenco is commercializing this new technology for liquefying natural gas. We began our development in 1994, dividing our program into two phases. The first phase is to develop a 500-gallon-per-day liquefier. This prototype, which has one engine and one refrigerator, is the first step toward a 500-gallon-a-day liquefier.

"This prototype now produces 100 gallons per day of LNG. It is 300 times more powerful than the original invention and has proven clearly that this technology works at large scale. Based on this success, we are now starting development of our 500-gallon-per-day prototype. Our efficiency target at that size is to liquefy 70% of the gas stream, while using the remaining 30% to power the burner.

"In the second phase of our program we will develop a 10,000-gallon-a-day liquefier, which will be similar to the 500-gallon-per-day liquefier, only larger. Both of these liquefiers will be small enough to install on transport trailers."

WOMAN NARRATOR: "Today most LNG is made in very large plants with capacities from 100,000 to several million gallons per day. The largest systems cost billions of dollars and take years to construct. The acoustic liquefier is much smaller, costs much less, and will be factory-built. However, it will achieve the same basic objective, to allow economic transport and storage of natural gas where pipelines don't exist.

"Small gas wells not connected to a gas pipeline could be brought into production with an acoustic liquefier. Offshore oil wells also produce some natural gas. An acoustic liquefier could withstand the hostile offshore environment and liquefy the gas for efficient storage and transport.

"Pipeline gas could be liquefied at LNG filling stations locally for fleet vehicles and near highways for trucks. The large quantities of natural gas generated by landfills could be collected, liquefied, and used as fuel for landfill vehicles.

"Local gas distribution companies could liquefy gas and store it when prices are low to have on hand when demand is high. Coal mines vent large amounts of gas over long periods of time. This wasted gas could be captured by acoustic liquefiers. Large-scale LNG storage systems have significant losses due to boiloff. A liquefier could recapture this boiloff."

DR. SWIFT: "High efficiency will be important for these applications. In parallel with Cryenco's commercialization efforts, Los Alamos is continuing research to push the efficiency of this technology higher and higher. Years ago, our first hardware was so inefficient it would have liquefied only 9% of the gas stream while burning the other 91%.

"When the prototype at Cryenco first liquefied 100 gallons per day, the efficiency was already 40% liquefied. The 500-gallon-a-day system should have an efficiency of about 70% liquefied. But even this is nowhere near the ultimate limit for efficiency for this technology. The acoustic liquefier is very new and our research has explored only a small fraction of all possible improvements to the thermodynamic and acoustic processes."

MR. WOLIAN: "With the help of Los Alamos we have made astounding progress. Cryenco is excited about bringing this new technology to the marketplace."

DR. KEOLIAN: Wasn't that perfect? [Computer demonstration PTR]

Remember, the orifice was dissipating energy and this was, I believe, driving Greg nuts. Here is the basic Stirling engine again. I put the elliptical indicator diagrams back on it. Notice there is a bunch of wave energy moving from left to right. The arrows are the heat coming in and out of the heat exchangers.

[Transparency 24]

Meanwhile, a fellow named Peter Ceperley, not too long ago, had noticed that in that diagram this phasing of displacing the gas, expanding it, displacing it, compressing it, has the displacement of the gas and the compression of the gas 90° out of phase.

If the displacement and the pressure are 90° out of phase, what is the phase relation between the velocity and the pressure?—180° or 0°—they are in phase. Acoustically, what does that mean, what kind of wave is that?—a traveling wave.

So Ceperley figured out that this is just an acoustic wave, a traveling acoustic wave through a regenerator. What if we can get an acoustic wave going around and around an annulus by putting one regenerator in that drives it (applying the necessary temperature gradients to the regenerator) and then use that traveling wave to make a refrigerator with another regenerator?

Dr. Arnott earlier brought up the problem of all that gas moving through the regenerator, through those little itty bitty pores. The viscosity of the gas is going to kill us. Ceperley made devices like those shown on this transparenncy and that is what happened. But he was on to the right idea.

[Transparency 25]

Back at Los Alamos these things were happening. There is the basic Stirling engine in (a). Then somebody else figured out, well, we can get rid of that piston on the right by putting in an orifice and a pulse tube, as I've described earlier, shown in (b). Then the Los Alamos guys had the idea, what if we could pull the energy out of the right-hand side and feed it back to the left-hand side acoustically?

For instance, in (c), we can get the phasing of the energy going back to the left side to be anything we want by adjusting the length of that transmission line, so we should be able to recover the energy that would otherwise be dissipated by the orifice.

Then they got more clever and cooked up the idea in (e), where they have this inductor and capacitor—they sort of messed around with the system here and published it in *Nature* (which I've passed out to you). In there is a description of this equivalent circuit. You can follow their math and it is not that hard to understand it that way.

[Transparency 26]

They built this thing and it looks like this. It has a couple of key inventions in it, key clever things. First of all, it is sort of like the Ceperley annulus in that energy is going to go around and around the loop at the end of the long tube, but this is not going to be an annular resonator of a wavelength around. It is actually just a small fraction of a wavelength.

They are using lumped elements. They want us to think of the slightly narrower part of the loop as an inductor or an inertance. The fatter end where the turn is made we should think of as a capacitor or compliance. Then we have the regenerator, a thermal buffer tube, and a junction. By playing with those L's, R's, and C's, as described in that paper, they found that they do not need a full wavelength going all the way around the loop.

But how do you get rid of that viscosity killing you in the regenerator? What you do is you realize that the acoustic power is the pressure swing times the volume velocity. We need that product to be high. The higher that is the more heat pumping we will get—remember that the motion that gave us the Stirling cycle heat pumping is the same as for a traveling acoustic wave going through a regenerator.

If we just use the ordinary traveling wave acoustic impedance ρc (ρa in our notation) to relate the pressure and velocity, our volume velocity is going to be too high. So what we need to do is up the pressure and decrease the volume velocity. What they did was, they stuck this loop gadget on the end of a big standing-wave tube. You can think of energy going around and around in the loop but riding on top of a giant pressure swing. They could have done it with a big piston coming in at the junction, where it says "To resonator" in the bottom figure, shoving gas in and out at just the right phasing.

[Transparency 27]

A picture of that thing looks like this.

An animation of it looks like this. [Computer demonstration]

You can study this at will, but I am running out of time here, so I am not going to describe the animation in too much detail.

DR. COSTLEY: How many wavelengths?

DR. KEOLIAN: It was on the last diagram, the length, there was a scale bar. It is about four meters long, altogether. The loop is a fraction of a meter.

This long tube, from end to end, is equivalent to a half-wavelength (it's actually more like a Helmholtz resonator to make it shorter). In the animation, you can think of the pressure swings and gas motion from the right of the junction, if you like, as being a piece of a big standing wave and think of the loop as being two branches coming off of the big standing wave.

[Transparency 28]

I am going to show you some diagrams now to help you understand this, and maybe it will help, as to what is essential to this thing. Here is what they are doing. They have a junction *J*, a slight constriction *L* for their inertance, compliance *C*, two pressure amplitudes *p+* and *p-* on either side of the regenerator, and a thermal buffer tube *TBT*. We will see that what is essential is that the pressure swings at *p+* are larger than and in phase with the pressure swings at *p-*.

Let's straighten the loop out as shown in the second drawing. Think of the loop as being the end of a big long standing-wave tube. The capacitor or compliance *C* is effectively the end of the tube. The inertance *L* is more or less additional length of the resonator tube. Here is our junction *J*, and here is our thermal buffer tube. The regenerator is effectively placed in a side branch to the main resonator tube. The end labeled *p+* is closer to the pressure antinode than the side labeled *p-*, and so the regenerator has a slightly larger pressure swing on one side than the other.

The impedance of the regenerator is basically that of a resistor because the pores are so small. And so the volume velocity through the regenerator is proportional to the pressure *difference* across it. But to have a large traveling wave, what we want is for the volume velocity to be in phase with the *average* of *p+* and *p-*, because the power flowing through the regenerator is the pressure *at* the regenerator (the average of its two ends) times the volume velocity ***through*** it (the difference of its two ends). So we want the difference in *p+* and *p-* to be in phase with their average. In the graphs at the bottom of the transparency we see that we can have a Δp, (which is proportional to the U) either by having $p+$ and $p-$ having slightly different amplitudes but being in phase or by $p+$ and $p-$ having the same amplitude but being slightly out of phase. In the first case U is in phase with the average pressure swings at the regenerator, which gives us Stirling heat pumping. In the second case we get U in quadrature with the pressure, so we end up with the viscous losses associated with the volume velocity going through the regenerator but without any beneficial heat pumping, which is bad.

If you think of $p+$ and $p-$ as being taken from different parts of a standing wave, as in the second diagram of the transparency, we can see that $p+$ is a bit larger than $p-$ and in phase with it, as we want. We can make the average p really, really big, because we are on the end of a big standing wave that could have a really high Q, so we can get lots of heat pumping without much viscous energy loss in the regenerator. That is really clever.

Another way of seeing it is through this third drawing. Imagine a Helmholtz resonator hiding inside a larger standing-wave tube. We have the compliance C being the volume of the little Helmholtz resonator, the inertance L is its neck, the junction J in front of that, and the thermal buffer tube, topologically, would be the space to the sides of the Helmholtz resonator. At the back end of the Helmholtz resonator we cut a hole and put in a regenerator.

There will be a slight Q amplification from this Helmholtz resonator and the pressure swings inside the resonator would be a little bit higher than the ones outside. But we want the $p+$ and $p-$ to be in phase. If there were a phase shift between them, as they are in the bottom right graph, we would get the Δp and that would be proportional to the U, but the U would be 90° out of phase with the p. We would be getting the viscous losses from the U going through the regenerator, but we would not be getting any heat pumping. For that we need the p in phase with the U.

So the right graph is bad phasing, and the left graph is good phasing.

[Transparency 29]

How do we get the good phasing? Think of the Helmholtz resonator like this. We tune it so that its resonant frequency is high compared to where we are operating it. We are on the wings of a resonance curve, so we get a little bit of pressure amplification, but without much phase shift if this is a high-Q resonator. That is the trick.

[Transparency 30]

Now I'm ready to conclude.

[Transparency 31]

Efficiency: This is the calculated performance from an older refrigerator that Swift's group built for Tektronics. It was a heat-driven standing-wave engine, a precursor to the TAD part of this TADOPTR I showed you in an animation. They had 1000 W work going to a load. They were using 5400 W of heater power. So the first law efficiency, "what you want" divided by "what you pay for," was about 18%.

The Carnot efficiency with the temperatures they had was 60%, so the percentage of Carnot they had was 30%. This was a well-designed engine—they got 30% of Carnot with a standing-wave device.

Where were the inefficiencies coming from? Half of it was being lost in the stack, that thermal relaxation loss we were talking about to get the darned phasing. They got away from that by going to Stirling. This got nuked.

There were viscous losses in the stack, 13%; 11% was being lost in the heat exchangers, all that gas moving back and forth. There was some heat leak to the room because they did not have enough insulation, regular losses on the walls of the resonator, some conduction along the stack and its casing, and losses from heat flowing through temperature differences in the heat exchangers.

But for the Stirling engine TASHE I just showed you, instead of 18% efficiency they got 30%—and they got 42% of the Carnot efficiency. This is good, this is going to be a good number. A pretty good commercial vapor compress unit gets 50% of Carnot. Some of the vapor compression chillers we are targeting that are in use by the Navy are less than this; the thermoacoustic Stirling beats them. An automobile engine is not this good. These guys are doing good on their first crack, so there is hope for high efficiency.

MR. APOSTOLOU: How good is....the one with the levers and so forth?

DR. KEOLIAN: Probably higher, I think, but it is more complicated. The acoustic one is with no moving parts.

[Transparency 32]

This transparency is a little small. It is from *The Physicists Desk Reference* from the American Physical Society. There is a chapter in there on energy usage. This is an interesting graph of where the energy in the country goes. These are in units of 10^{18} joules per year.

On the left-hand side is where the energy is coming from. There is nuclear power (a wee bit), hydro-power, a lot of it is burning natural gas, a lot of it is burning petroleum and natural gas liquids, and a lot of it is burning coal.

The middle and right of the diagram shows where the energy is going. A big hunk of it, more than a third of it, is used to generate electricity in this country; the rest is going somewhere else. Of the electricity, this wee bit down here is useful, but look how much is wasted in the generation and the transmission of it. That means if you run something on electricity in your home, you have to take into account that you have been warming a lot of bird feet on the way of getting the electricity to you. That makes heat-driven thermoacoustic engines look better than they might otherwise look.

Look how much of the fuel goes into household and commercial uses as just raw fuel usage. Presumably it is generating heat in homes and commercial buildings. That makes those heat pumps we were talking about perhaps a useful target application for thermoacoustics.

[Transparency 33]

Our weak spot in thermoacoustics, in general, is that our efficiency is not that great, compared to other heat engines. But Greg likes to make a point, which is, that Stirling has a better efficiency than vapor compression, yet you don't see any Stirling engines around. Why is that?

What we are plotting here is the capital cost of making various types of refrigerators, the initial cost of building it, versus its operating cost. The operating cost depends, let's say, loosely, on efficiency, so we have high efficiency on the vertical axis, which is the same thing as operating costs going down on the other vertical axis. On the bottom axis is how much money you paid up front to build the refrigerator.

Conventional piston driven Stirling has the highest ultimate efficiency, the little circle on the end of the line, but that is not very important, really, in general. Let's say you have only so much money you are willing to spend on a refrigerator and consider a vertical line, say, on the right-hand side of this graph. If you want to get really high efficiency, the best thing you could do is spend a lot of money and buy yourself a Stirling engine.

But if you are going to spend a medium amount of money for a refrigerator, say, draw a vertical in the middle of this graph. Then for that amount of money you can do better with a vapor compression engine than you can with a Stirling. The moral being that if you are willing to spend any amount of money to get the highest possible efficiency, then you will have to buy a Stirling engine, but most people do not do that. They have only so much money they are willing to spend, which is not an infinite, so they buy something else.

DR. DENARDO: You do not have maintenance costs in there.

DR. KEOLIAN: Yes, that is why operating cost is sort of loosely efficiency at this point. We push that the maintenance cost on the thermoacoustic refrigerators should also be low, so that also helps.

MS. PETCULESCU: The Stirling engines are still not in production.

DR. KEOLIAN: Yes, but they have been around a long, long time, the technology has been around for a hundred years.

MS. PETCULESCU: Aren't the costs a lot lower with mass production?

DR. KEOLIAN: Yes, that's so and not on here. This is a loose graph. But just bear with me to get the point I am trying to make.

Let's say a new technology like one of these thermoacoustic refrigerators comes along, it may not get to the same ultimate efficiency. The end of this new technology line might be below the vapor compression or the Stirling lines. But perhaps we can make thermoacoustics cheaper than vapor compression. Then if someone is willing to spend this amount of money on the left side of the graph, then the thermoacoustic version might be more desirable than vapor compression.

For another way of looking at it, consider the horizontals. If you want to get a certain efficiency, draw a horizontal line. At moderate efficiencies, you can get it for less capital cost with this new technology, whatever it is, than with vapor compression or traditional Stirling. It is only if you need the very highest efficiencies that these fancy technologies look better.

Our goal may not be so much to make high-efficiency thermoacoustic engines but cheap ones and then we can win. With that, I shall quit. Thank you.

MR. APOSTOLOU: Environmental issues led to this?

DR. KEOLIAN: Yes, I mentioned that in the beginning, I did not talk too much about it. We use helium, helium is nice. We do not use chlorinated flourocarbons (CFCs) or their replacements, the hydrogenated chloro-flourocarbons (HCFCs). The CFCs are banned now, and the HCFCs will be banned in about 30 years.

MR. APOSTOLOU: What is the alternative?

DR. KEOLIAN: There are not many alternatives, there are not that many choices. That is why we are doing this.

DR. COSTLEY: *[Helium has a high]*....speed of sound. It also has the long wavelength, but if you wanted something shorter, you could use —

DR. KEOLIAN: I had to skip that transparency (#18) for time. It turns out that the power of the engine does not depend on its length. It depends on its area. For a given length, if you changed gas, then you win by the higher speed-of-sound gas. Your operating frequency will go up but your power density will go up with helium. That is why we are driven toward helium.

Also, we are driven toward helium because we want a big temperature swing for a given compression, so we want all the noble gases. That also pushes us toward helium.

DR. COSTLEY: Say that last one, again.

DR. KEOLIAN: We want a large temperature swing for a given compression. That means we want a big γ, the ratio of specific heats Cp/Cv—the noble gases have the best of that. For performance we want a light noble gas, which is helium.

We can improve the Prandtl number by adding a little argon or xenon but there will be a hit on power density. If you want the most simple thing, use air.

Thank you.

POROUS MEDIA

James M. Sabatier
University of Mississippi

DR. SABATIER: I am going to talk about two things: acoustic seismic coupling, because that is kind of what I do and I was going to go through tubes and all this wonderful stuff, and then a little bit later on I was going to talk about mine detection. I have decided to turn it around and I am going to talk about mine detection first. That seems to be more exciting. I was really having a hard time staying awake this morning, because I have stayed up until midnight the past two or three nights in a row and waking up, as Roger Waxler would say, probably, at 4:30 every morning, but it is really 6:30.

[Slide]

Tom is here and Tom Muir made these slides for me in a motel room one time. We were going to go after lots of money. I came and told Tom how to find mines and I told Tom so well how to find mines that he was able to make slides that got me millions of dollars, so it is only appropriate that Tom is here so I can show him how well this technique really worked -- the advertising technique.

Let me tell you how we find a mine (Logan Hargrove taught me this). We take a big loudspeaker and broadcast at low frequency, 100 Hz, in the air, and we broadcast a tone, a swept tone or noise, and then you go to the spot on the ground and you measure the particle velocity of the surface of the ground. That is all you have to do.

If it is big, there is a mine. If it is not big, there is not a mine. That is what this cartoon is trying to say. Here is the mine. It can be an anti-personnel or an anti-tank mine. It is buried in the ground -- that is the surface of the ground. This is the particle velocity where there is no mine and this is the particle velocity where there is a mine. Tom was trying to tell me it is a lot bigger on top of the mine.

Here is the loudspeaker and the loudspeaker broadcasts 100 Hz. We have gone all the way to 10 kHz and 10 kHz is not good. Here, then, is how we detect the mine. We happen to be using a laser Doppler velocimeter, a device that you can go buy -- it is not quite at Walmart's yet, but in lots of places. It measures the velocity of the surface along the direction of the beam. It is a very easy idea; it is not complicated.

Anthony showed us on the first day that when the viscous penetration depth is large compared to the pore size, that is, when lamba or S, the shear wavenumber or this pore size parameter, is on the order of one, the sound speed is very low. It turns out it can be only 50 m/sec in the ground.

The sound comes in, in the air, at 340 m from our loudspeaker. It hits the ground and it has to go 50, so Snell says it has to bend toward the normal, so it bends strongly toward the normal. At low frequencies, on Anthony's curve, or as the frequency increases, the phase velocity of the wave in the tube increases. At higher frequencies it will not bend as much toward the ground, but it never gets equal to the speed of sound in air. It always stays less than the speed of sound.

This is a patch of the ground, say, 1 m by 1 m. This is where the mine is buried and this is the light coming in. It looks right here but in a really steady position. It is bad if the laser beam moves a lot. It measures the velocity, then kind of steps across like this and the laser beam makes a raster scan, a digital raster scan. It stops and collects and measures the velocity.

[Slide]

This is a picture of, I think, an M19 mine. It is an anti-tank mine. It is 12 inches across, it is square, too. It has 21 pounds of TNT in it and it is made for taking out a tank. This is an image of the RMS particle velocity on the surface of the ground on a 1m by 1m area where this mine is buried at a single tone, at one frequency.

When you do that measurement, you take the time-domain signal in, you do the Fourier transform, you go look at the instantaneous FFT, the real part of it, you see this phenomenon -- I apologize, it is probably the magnitude, it is the magnitude of the FFT. The mine is right here.

This is the same thing, the same spot on the ground, except now, since we broadcast -- we trigger -- the signal and we trigger the A to D when we send out the band of noise. We get the velocity at every point and we know the phase of the velocity at every point relative to all the other points.

This is an image of what we might call the instantaneous velocity at a single frequency on the surface of the ground. The frequency is around 100 Hz. This is 1 m by 1 m and this is the velocity, so the ground goes up and the ground goes down. You kind of expect that, because the wavelength in air is 3.5 m at 100 Hz. This is only a fraction of half of a wavelength, so it pushes the ground down and pulls it back up and pushes it down.

Where the mine is, it is out of phase. These two phenomena happen only in narrow-band frequencies and it does not happen at every frequency.

[Slide]

That is not how it started, that is not the way our work started. It started in about 1986, I guess. This is frequency from 0 to 500 Hz. This is the magnitude of the velocity that we measure with a geophone. A geophone is something like an accelerometer but it is an integral of it. This is a device that measures velocity. It is a mass on a spring with a coil of wire around a magnet. The mass is a magnet, so when it moves up and down you get a voltage that is proportional to velocity. It has all been normalized. It goes to two.

This is the velocity away from this mine measured with the geophone. All I did was set the geophone on the ground. It is an M19 case, it is empty. I filled it with plaster of Paris. It is buried in the dirt behind the old acoustics lab at NCPA. I lost it, by the way, and I never could find it. I could not figure out where it was. We moved from the building before we got back into the business. This is the velocity measurement with the geophone and that is on the target. This is on the target and this is the background. Hank Bass said, "Jim, just integrate those two curves and get a number and make a map."

About this time Ken Gilbert suggested we should use a vibrometer, a laser Doppler vibrometer to do this, and Hank suggested we ought to write a patent. So we did. Ken and I drafted this -- I think Ken mostly wrote it, I do not know, it has been too long to remember, but we wrote this patent and we did whatever you do and it got classified.

The contract we had was broken, because we did not have a facility clearance at the University. I had a clearance and Ken, everybody had clearance, but we did not have a facility clearance. The security officer came, showed up, and took everything, including that graph, and kind of put us out of business, and then they gave a million dollars to another company to see if they could do this with a vibrometer. Needless to say, when they called me and asked me for help, you can guess what I told them. (laughter)

In about 1994, right after the Persian Gulf war, we started getting a lot of press about mines, because doctors were getting out of jeeps and stepping on mines, so Congress started money in and more funding became available. We got money from ONR at NCPA to look at that high-frequency work, because we wanted to see if we could increase the frequency.

If you work at 100 Hz, you have to wait for this 100-Hz cycle and that is slow. The guys who do this with radar work up around a gigahertz, so they have to wait only 1 gHz of a cycle to get their data, so they can go a lot faster.

[Slide]

These are measurements with the geophone. This is a SIM30. It is 30 inches in diameter and it is buried 3 inches, and it is buried in the dirt behind Angsells [phonetic], we have a little plot back there. It is highly absorptive soil to this slow wave that we are talking about. The grain size is probably only 30 or 40 microns of the soil. The pore size is about a third of that. The sound does not go very far. This works better than we expected.

The mine is buried at maybe zero centimeters and zero corresponds to the center of the mine. It is 3 inches deep. The mine is right here and we put the geophone on top of the mine at zero. We use a swept sine from 50 to 200 Hz and we look at the velocity, the magnitude of the geophone. As you go up, it gets big. Then we look at + and -2, so those are the three red curves (I do not know which ones are which). This is + and -4 inches -- that is the black. This is + and -6 inches down here, both of the blue ones. Green is + or 8, and 10 is brown. So over the mine there is a great big signal and it cuts off very, very fast -- that is with the geophone.

Here is a piece of wood at 5 inches. It is a 10-inch plywood disk that we painted red and that is what happens when you look at that piece of wood.

This is a concrete paver. It is about 10 or 12 inches in diameter, it is circular. You put these on your patio to walk on. That is the signal you get at a concrete paver at 8 inches.This one had gravel on the top of it. I painted it because I was trying to seal the pores. This one had gravel on the top, this is one of those fancy patio pavers.

This is another concrete disk, it is 8 inches. On this one there was no gravel on top of it. I like to believe that this one is not showing up because it is porous concrete, but I do not really believe that. I think there are mines that we cannot see, that is what I really believe.

We did all this. We received our current funding in September of 1997. That summer the sponsor from Fort Belvoir visited us. We bought this big LTV that you saw. He said, "Jim, this is really nice work you're doing, but you need to come to A.P. Hill and do a demo at A.P. Hill, where we have real mines buried in the ground. My work is what is called 6.2. I have to do a demo, that is the whole purpose of the money, is for me to go do a demo and show how well it works.

He said, "But there are a lot of political ramifications, don't come over here and fail." That is what he told me and that was all I needed to hear.

[Slide]

In November we took off. We knew something about an LTV. This is Pat Arnott's slide, it came from a paper by Pat and me. When we applied for the patent, I guess I did what I was told, but all I said I wanted was a vibrometer instead of a geophone, so somebody got me a vibrometer, and this is it right here. It was made by Dentech and we got this in 1987.

Pat took it outdoors to see if he could measure what a geophone measured, not find mines, that is all classified now. We wanted to see if a vibrometer does what a geophone does. I showed we could find mines with the geophone, or buried objects, so now could we do it with this vibrometer.

Pat had to bring an optics table out into the field behind Hank's house, and then he had to have additional isolation right here. Then he had to put a big box on top of this thing to try to shield it from the wind. He buried a geophone in the ground, flush, and he reflected the laser beam off of a mirror and down to this geophone.

He did this in the lab with the geophone on the shaker. He mounted the geophone on a shaker and repeated this experiment and computed the transfer function, the voltage from the geophone, change it to velocity, divide it by the voltage from the vibrometer, change to velocity, and it is exactly 1, or very, very near 1. So he wanted to take it outdoors to see what happens outdoors.

[Slide]

To use it for a source he had an explosion. We had a small propane cannon that he put off at about 100 m and that thing was going to set off an explosion and the pulse was going to come by and the geophone would respond to that and the LDV would respond to that and he was going to compare those signals. That is the time-domain trace.

This is the velocity, 150 microns per second, plus or minus, going on for about 200 msec. One of these is the LDV and one of them is the geophone. Time-domain traces do not always tell you a whole lot, so he used a loudspeaker to do the same experiment.

The dashed line at zero is the lab. This is the transfer function between the geophone and the LDV in the laboratory. One of these curves, and I no longer remember which one, is the transfer of function when the LDV is looking at the buried geophone and the loudspeakers are sweeping from 100 to 500 Hz. We would like for that to be zero dB, but it is not, it is one of these lines. They start to deviate pretty badly.

The other one is when you just take the laser and instead of looking at the top of the geophone you look at the ground next to it, and I suspect it is the one that is the worst, whichever one that might be. My guess is it is this dashed dotted line.

In any case, these data show that it works, that you can use an LDV outdoors to measure the vibrations on the ground and you can get the same thing you get with the geophone, so you ought to be able to find mines, and I knew that.

[Slide]

We did not have time to do much work with this vibrometer when we got it. We received it in April of 1997. I go off in June, July, and most of August to work with high school physics teachers and I had some young people who were going to try to use it to find mines. I got back in September and they said, "It doesn't work, we can't find anything buried." I said, "Well, I'll go with you to the field and help you." We got out there and it was just a matter of -- I do not know what it was a matter of, but it started working (kind of voodoo).

We went to A.P. Hill in November of 1998. Fort A.P. Hill is in Virginia, near Fredricksburg, and they had these lanes where they buried mines. This is what they call a dirt road. I told them they should not call it a dirt road, they should call it something else. A gravel road is what it really is, it has lots of gravel in it, there is a rock right there. If you take your pocket knife and you push you pocket knife in the ground, you push about a centimeter and you hit a rock. The rocks are all about an inch in size and they are packed in clay and sand.

In the background is another road and they call that the blue gravel road and it is gray. That one has gravel all over the surface, about 1-inch sized gravel, and there is a layer of gravel on the surface.

We got out there, and here is the vibrometer. Here is the truck with some electronics in it, there is a cable, there is a loudspeaker. We were looking at anti-personnel mines, which are a lot smaller. I just believed you had to have shorter wavelengths to find anti-personnel mines, so that is what that horn is about. That thing does not do much at 100 Hz but around 500 Hz it gets going pretty good, or 300 Hz, maybe. There is the loudspeaker, there is a mine, they are looking at the mine right there.

The first day, this guy (he was the sponsor) and about three or four of his buddies came down to watch. We were trying to hook cables up and they were looking over our shoulders and we were so nervous that I could hardly put a B&C cable onto the B&C connector.

This guy right here, he really gets nervous. He makes me look like a cool dude -- this was Ron Craig. Ron was shaking so much I told all the guys from Belvoir, "Do you have a coffee pot, let's go get some coffee," and we left. When we came back, they had everything set up and it was working, so we started scanning mines.

[Slide]

I have kind of explained this already, but the vibrometer measures the velocity of the ground and it does it at pixel spots on the ground. We get the magnitude of that and we look at it in bands and often we normalize. Then this instantaneous velocity, which is that animation I showed you, we know the phase, so you can put the phase in here, and we know the magnitude, and you can do some kind of animation.

[Slide]

Here are measurements now with the vibrometer. This is in micrometers per second, microns per second. It goes to 30. These are spots on the ground on that clay gravel road, the "dirt road," where there is no target. There are a half-dozen spots here that are all within a square meter and they all look very similar from 75 to 275 Hz. We broadcast a band of noise from 75 to 275 Hz and then we measure the velocity, we take the Fourier transform, and that is the Fourier transform, that is the magnitude of the Fourier transform.

Here are spots down the length of the gravel road from 3 m all the way down to 15 m and they pretty much all look the same, too.

Now let me tell you how these lanes are made, the Army Corps made these. They came in and they excavated a foot of soil and they hauled it away and they put this fill material that they make roads with. If you are engineer you have to make roads in some prescribed manner. They put in 4 inches of this material and they packed it, put another 4 inches and packed it, and they made it 3-m wide and it was crowned, but it is very homogeneous, and that is what these data show. If you look around, everything is the same.

[Slide]

These are now spots on top of a mine. This one here is something called an electromagnetic 12, EM12. It is a simulant for radar techniques. It is buried at an inch. The black is the off-target spectrum in micrometers per second (this goes to 40 now). This is the velocity at one point with the laser looking on the center of the mine, so it is highly elevated.

This is something called a TM62P made by the Russians; P is plastic. They have an M version that is easy to find with radar. It is buried 2 inches and you see a similar phenomenon as that one.

This is an Italian mine. By the way, these are all anti-tank mines, a foot in size. This one is about 10 inches. It is the most difficult mine to detect of the anti-tank mines for anybody. It is buried one inch and this is the on-target velocity. That is the off-target velocity.

This is an M19, which is that big anti-tank mine, at 5 inches. That is the on target. This is the off target, black is off target.

I should tell you that when we first got there they told me not to walk on the lanes. I thought that was a good idea, you should not walk on the lanes. But then I learned that there was no detonator in these things and something called the booster was taken out and it was filled with some material that makes it look okay for radar. Usually it is filled with calk, rubber.

Later on I learned that it was okay to walk on them, because I saw a pickup truck driving down the lanes. (laughter)

[Slide]

Here are some results of scanning VS. This is the brother of the 1.6, it is a 2.2. It is about 10 inches in diameter and has a nice little carrying handle right here. It is made by the Italians and it says it has 4.7 pounds of "something" in it (I do not know what it is). It is an anti-tank mine.

These are spots where we scanned. That is the ground right there, this is a video image of the ground. This is a very nice vibrometer. This white circle we added. The color of the dot is the area under the curve on the previous graphs, with these numbers assigned to them. (There are some problems with the units I do not really quite understand.)

In any case, if it is green, in Mississippi there "ain't" no mines. If it is red, there is something there. You can smooth, you can interpolate between these things so you make this a nice smooth curve, and that is how I made those nice red and green circles that you saw -- here is one right here. You can see an effect out to here, yhu can see these darker green dots out there.

We made pictures like the previous slide for about 40 mines while we were there on the first trip, and the sponsor could not believe it, because everything is green all the time. If there is a mine, it is green, if there is no mine it is green. If there is a mine, it shows up as red circle.

So he took us off to a road in the woods. This is an image of that road and you can see leaves on the surface right here. Everything that is not this kind of light-colored stuff right here is bare soil, all this other stuff is short grass that was growing -- it was November, so you can see leaves, and there are leaves on the ground right here (I suspect these spots right here are leaves).

I asked, "What is this lane for?" He said, "Well, this is where we started about five years ago and nobody could find any mines on this lane, so we built the other two lanes that you've been working on."

This one, if it were not for the grass, is easier, if you could just get the grass off the surface but leave the roots, it would be an easier spot. This is the mine that is buried. That is its image when it is 3 inches in this so-called off-road site.

[Slide]

We looked at anti-personnel mines.We did not plan to do this. This is a PMA3. This one, for some reason, really has explosives in it. This is all plastic. There is the signal from 350 Hz to 700 Hz. Again, red is on the target, black is off. That is in that clay surface, the so-called dirt road, at 2 inches.

This is 1 inch in the granite gravel road. It is just loose granite on the surface. We do not look in these frequencies any more for these mines, because there is a lot of clutter in high frequencies, you see a lot of elevated signals.

[Slide]

This is a wooden mine. This is 4 inches by 8 inches, 1 inch deep and 2 inches deep, and these are images of them, or those are the spectra. This is running around 5 or 10 and the signal here is up around 20 μm/sec. In every case the velocity is increased.

[Slide]

This is a U.S.-made M14, an anti-personnel mine. It is 2 inches in diameter. It is the hardest mine to detect of the anti-personnel mines and the U.S. puts them everywhere, apparently.

This is an Italian-made mine and it is 2 inches and this is this clay gravel road. You can see two strings right here. They told us the mine was between the two strings. If you notice, there is a rock right here holding the strings back, because the first time we scanned there was nothing there.

I said, "It must not be here, so maybe it is off." We pulled it back and we slid this image over a little bit and we scanned again. This is only about 10 inches right here. The size of these red circles is the size of the mine, approximately. The positional accuracy of finding the mine is about an inch, that is how accurate it is in terms of locating it. With other systems it is sometimes a meter, which is not unreasonable for locating these things.

[Slide]

During that first trip in November of 1998 we scanned all of these kinds of anti-tank mines in three types of roads and the off road as well as all these anti-personnel mines. These numbers are the depths in inches and we made an image of every one of these, as we showed.

[Slide]

After we got back, I traveled to Fort Belvoir and gave a presentation and there were a lot of skeptics, and they probably should be. We showed them these data back at the Countermine Division at A.P. Hill and they were duly impressed, but someone said, "You have to do a blind test. If you know there is a mine in the ground and you're looking right there, you can always find it." That is what they believed and, for the most part, they are kind of right, but sometimes you can know there is a mine there and you cannot see it. Those are the facts.

They asked us to come back and do a blind test. This is April. You will notice there is a different-colored forklift. There is a story about this forklift. This thing is slow, I did not tell you that, this is really, really slow. It takes us, on a good day, to scan this square meter, about 3 minutes, maybe 2½ minutes. On a bad day it takes 20 minutes.

Every time we wanted to move to another spot, we had to call this guy on a radio, the forklift operator. He would come down and start the forklift and move it 3 feet. I watched him, and I climbed up there and said, "I can do that, let me drive it." He said, "You can't, sorry."

The next time I rented my own forklift, put it on my University American Express card. I got over there and Dan Costley, I think, was with me. They were in a tither, "You can't drive that forklift." I drove it off the truck. When I got on the truck, the guy unhooked his truck and said, "It's yours," and gave me the key. I asked, "You don't drive it off?" He said, "Nope, I just deliver."

I got in and could not get the key in, so he showed he how to do that. (laughter)

When I turned the key, nothing happened, so he told me to put it in neutral. I turned the key and nothing happened. He told me to put on the brake and I put on the brake. It started. So then I put in forward and it would not go, so he told me to take brake off. (laughter)

I get it off and I cannot get out, could not open the door, so I drive it to this and I set it up near here, and the guys at A.P. Hill -- the engineers, now -- were talking with the physicists and they were trying to figure out what they were going to do. They came over and said, "We maybe think you shouldn't be driving this forklift."

I said, "You're probably right but the contract says I'm the only one who can drive it, so that's the way it's going to have to be," so I drove the forklift and I drive it every time.

We changed things a little bit. One of the problems is, here are the loudspeakers. They are 18-inch drivers. We started out with one, then we went to two, then we went to four, trying to find the mines we cannot see by getting more and more loudspeakers (it does not work very well, actually).

This is the spot where we are going to scan. That is the vibrometer. The sound level right here is not a whole different from the sound level up there. In fact, what we are trying to do is to get this device far away from the drivers compared to the ground from the drivers; we are trying to take advantage of spreading the farther you are from the drivers. The ground does not vibrate a whole lot more than the platform, and this is an interferometer, it does not care who is moving; it measures the relative position between the two surfaces as a function of time. We had this thing very, very high in the air, up to maybe 7 or 8 m, trying to get farther away.

They wanted us to do this blind test and what they told us was, "We're going to give you 60 spots, we're going to put 60 golf tees down, 30 on the dirt road and 30 on the blue gravel road and you go see what happens and come back and tell us what your score is."

The rule is that the mine is going to be somewhere within a square meter, the center of the mine, so the center of the mine could have been on the apex, the corner of this square meter, and if we scanned a square meter we would see only 25% of the mine, so we scanned 1.5 by 1.5 m. Actually, there is an interesting story about that and I will get to it in a second.

[Slide]

We thought we should practice before the blind test started. This was in the middle of the week and this is an M15 at 3 inches on the blue gravel road, 1.5 m by 1.5 m. We know the mine is there, they is why it is in the middle. This is the velocity going up to 60 μm/sec, so that is the highest number right there. The frequency band in which this was being imaged is between 120 and 150 Hz.

We did this for a while and we went back to the motel on Saturday -- we had collected quite a bit of data like this on Friday -- and we started worrying about this blind test. We had never really thought about what a blind test might be like, so we thought about it over dinner and got intimidated, because we said there were going to be some spots without mines and we had not scanned a lot of spots without mines and what are we going to do if there is no mine? Well, we are going to look deep into the background in the noise and see if we can see this elevated signal.

We came back on Saturday and I closed up the windows on the truck and closed the door and one guy stayed in and the other two got out and we moved the LDV around and we would not tell him where we were looking. We would put no mine or we would put the mine half on the circle or we would put the mine right in the middle of it and every time he got it right. We collected a lot of data like that.

That weekend we renamed all the files, because our coding system tells us the name of the mine, so we renamed all the files so we would not know there was a mine there and we gave ourselves a blind test on the computer. We learned some stuff by doing that. We learned that we could miss some mines, that when we had a spot that did not have a mine and we looked really hard, we could find things. It never looked like this, so we could be fooled.

[Slide]

So we came up with a strategy over that weekend and this was our strategy. We collected quite a bit of data and were able to do this. We said we were going to scan 1.5 by 1.5 m with a 10-cm pixel size spatial resolution and we were going to use a 10-Hz frequency resolution on our FFT and we were going to do three averages. That takes something like two minutes.

We decided that if we saw a mine we would get on with business. If we did not see a mine we would rescan with 7-cm resolution and 5-Hz frequency resolution at closer points together; we were going to get more points on top of the mine if there is one there. We were going to do five averages instead of three, and that takes a little bit longer. We had lots of data to justify this.

We had never done this before for anti-tank mines. We said, well, maybe we might want to look really close in an area and we invented this little thing in the motel room, half-meter by a half-meter, with a 3-cm pixel size or spatial resolution, and this was going to take five minutes. That is what we did.

[Slide]

We go out the next morning and we are getting set up. I did not tell you this, but when the wind would blow the thing does not work and the wind always blows. We had learned that if we bought a U-Haul cardboard box and put it on top of the vibrometer and duct-taped it down, it worked a heck of a lot better.

When I cut the cardboard box I cut through one of those cables that ran all the mirrors and we were really nervous, so this was a good thing. We knew how to solder and the wires happened to be all color-coded, so we rebuilt it. By the time we were through working, all our nerves were gone.

That is a scan in the blind test, it was not the first one, so we said that must be a mine, and they were not happy with our telling them that. They wanted us to give it a score, 1, 2, 3, or 4. One is a mine, definitely a mine.

Four is definitely not a mine. Two is maybe that is a mine. Three is maybe it is not a mine. I did not know the damned difference, I really did not, but we scored this a one.

[Slide]

Here is another one. I do not know what they are. In fact, my sponsor does not know; it is kind of a double-blind test. The people out there do not know where they are, but there are some surveyors who come from IDA, Institute for Defense Analysis, and a judge. The judge sits there and writes down what you say.

Here is a spot where there is no mine, that goes to 70 and that goes to 70 μm/sec.

[Slide]

Here is one of the coarse scans. We get this a lot. We are close to the loudspeakers and although I do not always understand it, at some frequencies you get bands of red across here of high velocity and we are used to that and we can live with that.

But this spot right here got us interested. We though maybe that might be a really deep mine. It is around 40 μm/sec. It is in high frequency, pretty high frequency compared to the other mines I showed you before.

[Slide]

We rescanned it with 7 cm. Now it has been auto-scaled, so this is still only 40, so maybe that is a mine.

[Slide]

We decided to do this zoom scan on it. That is now 50 by 50 and now you see that it looks like three things and we ruled it as not a mine.

When people walk, you can see here this is someone's heel, this print right here, it went through the sand. These are car tracks here in the sand. These things often show up as high velocities. If you have enough area, then you can see what that is and you can tell that is not a mine.

Most of the time the mines are great big red circles, it is just the ones that are deep that you have to worry about, and maybe we do not know that we should not be worried yet, but we really worry about that, and you will see why as I go on.

[Slide]

In the blind test this was the target set; there were supposed to be 60 spots, but there are only 59. Nineteen of them contain mines, they are all anti-tank mines, and these are depths they were at and you can see that most of them were up here in the shallow depths.

[Slide]

We did not know this. Of the remaining ones that did not have mines, 31 of them were just spots down the lane were they knew there was no mine, but nine of them were kind of special, nine of the spots without mines. These lanes were built for advanced mine detectors, something that goes on a jeep and you drive it down the road at 10 km per hour, so there are a lot of ground-penetrating radar mine-type data and it has GPS coordinates and they get lots of false alarms.

These nine spots, everyone who has a ground-penetrating radar and drives it down the road on these nine spots all say it is a false alarm, so they call it contractor-to-contractor false alarms, and those were thrown in and we did not know that.

[Slide]

These are the results. We missed one of the mines. That is a 95% probability of detection. It has never been done before. The best anybody else had ever done was about 80%.

The mine that we missed -- this is from the people at IDA and this is their quotation -- "The mine at 6 inches missed but it is clearly present in the data." It goes on to say that the operator said it was a big rock.

I am not going to show you this, but they showed me lots of pictures of rocks, photographs, from rocks this big all the way down to rocks this big, all sizes. We just knew they had to give us a rock, so we thought we could tell how big the mines were by counting those little red dots across and multiplying by the pixel size, we could get the diameter, so we had something that was 50 cm, and we said that was too big to be a mine.

Subsequently, we found out that a lot of them are 50 cm. It was a big red circle and we thought it was too big, a mistake. We have not seen a rock at this time in our measurements.

In 41 patches we found a false alarm, only one, and, most importantly, in the non-cluttered spots, none of them sounded an alarm, so this was very promising for people who are in the mine-hunting business, because radar-type devices can find most of the mines but they find immense numbers of false alarms.

This was the false-alarm rate. They often find numbers that are 10 times this big, one mine every two square meters. You have to get this number a lot farther down or you spend all your time looking for land mines.

[Slide]

They still did not believe us, so they sent us to Yuma to look for rocks. The soil at Yuma is like powder, talcum powder. It is very, very fine-grained stuff. They built four lanes, put in a zillion mines, all anti-tank mines, buried them at 5, 10, and 15 cm. They had three types.

The soil really turns to talcum powder, so they went out with a water truck and they wet it down with water. When you do that, this desert soil at Yuma turns to rock, it gets really, really hard. You can drive a forklift across and it does not make a track in the soil, that is how hard it is.

They did that on one of the lanes. They were going to do it with all of them, but there was another group there from Stamford International who had a radar and they could not see any of the mines in the one that they had wet down three days before, so they did not wet any of the other lanes down.

We found 110 mines, apparently there were 18 that were not visible, and we did not find any rocks. We did not see any false alarms -- we saw one false alarm, and there we could dig. We are not allowed to dig at this Fort A.P. Hill place. These guys were a lot more relaxed.

We went back and we found a big red circle, so we went back and we dug and it was a pile of debris, stems and branches from a desert plant. When they made the roads, a road grader scraped this stuff up and it just happened to be -- first, I thought it was the root mass from a plant, and when the grader had sheared off the top of the plant, it was just the root mass. I could see how that might be a mine, but it was not. I do not really understand how it got there like that, but it did. We did see any rocks and there were a lot of rocks, a lot of rocks the size of your hand.

Remember, we are looking about 6 and 7 cm apart and looking for things about that big. They do not show up in the low frequency.

We did a blind demo. There were 27 patches. We found 93% of the mines, but we had a false-alarm rate of 0.11, but there were some holes, they dug some round holes in there. I screamed foul. I think that if they did a radar test and buried metal plates, people would scream foul.

There is another system out, called nuclear quadripole resonance (now it is just quadripole resonance), and if you bury chunks of TNT, they would scream foul, too, because it is not a mine, it is just a chunk of TNT.

But in any case, we will always find holes, and you will see why as we go on.

[Slide]

Here is a rock -- that is a 2.5-inch tape measure there. We dig a hole and bury that rock in it, someone thinks it is about that deep, it does not look like a round mine.

[Slide]

Here is the image. This is the velocity from 100 to 160 Hz, this is where we can see it. The red and the blue are off the target, so the velocity is less than 5 microns per second, but on the target it goes up a little bit. We often look down here to see if we can find mines.

[Slide]

There is an image of it, so we would probably miss that. There are some times where you see things like this and it gets you in trouble.

I think we should take a break, so we will stop for 10 minutes, then we will come back.

DR. SABATIER: Albert Migliori asked me a question.

DR. MIGLIORI: With all this success now, you have no possibility of ever spending enough time to understand what is going on? (laughter)

DR. SABATIER: That is what I was going to try to do. I do not understand all that is going on and that, hopefully, will come through.

Let me talk about the rock first. I showed that picture of the rock and Albert said, "Well, do you see rocks or you don't see rocks?" It is not an easy answer, because I buried that rock. If you go to a soil and you dig a hole, you know you cannot put all the soil back in. That is a rule of thumb. Even if you save all the soil, it does not all go back in.

That will generate an increased velocity if you dig a hole, so I do not know if that slightly increased velocity is due to the fact that we disturbed the soil or if it is due to rock. We are going to do some experiments -- there are obviously lots of experiments that you would be able to do and maybe I will describe them at the end when you understand a little more about the physics.

MS. HIGHTOWER: Can you tell from your data how deep they are, the mines are?

DR. SABATIER: Not really. If the porous material properties were homogeneous, then we could. You will see some things like that.

DR. STERN: If you are looking for mines, this should not happen. I mean, everybody dug a hole, somebody dug a hole. I guess the real question is how long does it take for a hole to recover so that now you cannot tell whether it is a week or if it is a month.

DR. SABATIER: The mine that I showed you that was in the off-road lane had been in the ground for three years when we looked at it, and there are some mines in that lane that we do not detect.

MR. APOSTOLOU: ....to the loudspeakers, how loud is it?

DR. SABATIER: Do you work for OSHA? (laughter)

If you say yes, I would say it is 80 dBA where the operator sits, because if you look in Kinsler and Frye, they have a chapter on environmental noise and it says you can listen to 80 dBA all day long, no hearing protection is required.

We get the sound level where the spot is we are scanning in the band we are showing you to about 120 linear C-weighted.

DR. GILBERT: Instead of digging a hole, has anyone tried just covering it up with dirt so that there is no difference in the character of the soil compared to the surroundings? Cover the thing up with 6 inches of dirt?

DR. SABATIER: Just to put a target on top of the ground and just put dirt on top of it? We do this in sand, clean washed homogeneous sand. It works great and you can take the mine back out and smooth the sand out and it does not show up.

But what happens in a soil -- this is leaving porous material and talking about granular materials, which is a subtle point but may be important. Soil is in some what we might call a strained state and it is due to whatever its history has been.

If you take the soil and put it in what is called an odiometer and you increase the pressure, the stress, and you measure the strain, the strain does not change until you get the stress back up to its "remembered" stress. Then, when you go beyond that point, the grains start to move, they start to slip, and it starts to strain again.

Now, if you release the pressure and you start turning the pressure back up and you watch the strain, the strain does not start to change until you get back to that. The acoustics is the same way. We have done an experiment where we measured compressional and shear wave velocities in a soil and we stress them and we measure the strain and we measure the compressional wave velocity as we do that, then we relax the stress and measure the compressional wave velocity and we get the same number. Until you get the stress back up to where it was and start to strain again, the velocity does not increase until that point.

DR. GILBERT: But do you see rocks or not? You have told me that you do see rocks.

DR. SABATIER: I showed you a rock that I saw.

DR. GILBERT: But then you said maybe it was because the soil was --

DR. SABATIER: In that case. What I think we see are right circular cylinders that are the appropriate size, concrete disks. That rock was not a concrete disk, there is scattering that is going on.

MR. PETCULESCU: Under dense grass, does grass affect the vibrometer?

DR. SABATIER: Of course, but it is not going to affect the acoustics, but it will certainly affect the vibrometer, so I do not say that maybe this vibrometer is the instrument that you use to go find mines. It is an instrument that could be used in some cases, but there are other techniques. Albert and other people were saying you could use radar. Radar goes right through gas and hits the surface of the ground, so if you could make a Doppler radar with enough sensitivity and the right pixel spatial resolution, that would work.

Realize, also, if the mines are really on the surface and they are covered with grass, probably just an impedance tool would be a good way to find it.

DR. CRUM: Since Doppler radar is a different mechanism, if you combined the two you could reduce the false-alarm rate.

DR. SABATIER: I did not say it, but we are working with another company that has synthetic aperture radar and we have always scanned the same spots in the last two years, we always have done this simultaneously, and our measurements are orthogonal in terms of false alarms, so there is some potential there, and their spot size is about the same as ours, so they are looking at 5 to 7 cm. The data could be fused down at that fundamental level, which is important.

DR. CRUM: They can scan a lot faster, so therefore, if they could say they got a lot of false alarms, and you could go in and check out the false alarms, then --

DR. SABATIER: That is one scenario and they have a name for this: They call it a confirmatory sensor. You drive down the road and they mark all the spots -- they have something called an overpass vehicle -- then you come back with the slower technique and rule out the false alarms, so speed does not have to be that important.

We are not going to fight any wars where we have to breach mine fields any more. That is the old view of how we fight wars. We demonstrated recently that we do not fight wars anyway, so we do not have to go fast. What we really need to do is clean up the problem. Maybe I should say a little bit about this.

An anti-personnel mine is a lot harder for us. We are scheduled to do one of these blind tests on anti-personnel mines as soon as it cools off in Virginia. We tried the high frequencies, around 400 to 700 Hz. When we do that, we see lots of clutter, footprints. The heel of a footprint on a dirt road, on a sandy road, if it crusts the sand, it shows up as a high velocity, but if we drop it down (and we just recently learned this) and work between, say, 175 to 250 Hz, those things do not show up but the anti-personnel mines do, but not the little tiny one, the U.S. one that is less than 2 inches. We will have to give up on that one, but there will not be any false alarms.

The wisdom of most people I talk to who are in the mine-hunting business believe you want low false-alarm rates, because that is what they think slows people down. That is probably artificial, because someone says you have to go 5 km per hour. No one ever decided how fast you could go. Someone just said this is how fast you have to go, and that does not make a lot of sense.

DR. CRUM: Wouldn't a false negative be more important than a false negative?

DR. SABATIER: Sure.

Someone once told me there was on the order of 100 million mines in the ground; that is 10^8, and the current cost is about 10^3 dollars to remove a mine, so that is 10^{11} dollars. That is a big industry.

Now I am going to talk not about mines but just the coupling of sound into the ground. People have done -- actually, mostly us (for some reason nobody else really cares about this problem), more people now are caring now that you can find land mines but in the old days nobody really cared about this. When you go to ASA meetings, I am the only person who ever talks about this. It is just kind of the way it is.

[Slide]

Here is a loudspeaker. Here is the surface of the ground. Here is the geophone and usually you bury these things. You bury the geophone, and I put a reference microphone right here and maybe you are worried about this angle. You measure the frequency response of the geophone divided by the microphone.

Other people are interested in ground impedance -- Ken Gilbert talked about that. This is really outdoor sound propagation and people want to know what the ground impedance is, and that is how we got into this business. Before they knew about turbulence, ground impedance was really important, but now that they know about turbulence nobody cares about ground impedance.

You can imagine having two microphones. I call this a reference microphone and the combination of these two the gradient microphone, so that gives you <u>delta</u> P/<u>delta</u> V. That is impedance. When people measure ground impedance they do not do it this way. You could, you could use a two-microphone impedance, too. Does anybody know what that is?

Tim knows what it is, Pat knows what it is. They built one. You can buy them from B&K. There is a standard for measuring ground impedance with these devices. You essentially have a tube here and you pound this tube in the ground and you measure this in the tube and you get impedance. It is notoriously a very hard measurement to make.

We bury a microphone. We put a microphone in a brass tube and we push it in the ground. In the ground we measure two things. We think this measures velocity and we think that measures pressure in a porous material.

[Slide]

These are some data taken by Hank Bass, 150 Hz to 1500 Hz. This is zero. This is the real part of the impedance and it has been divided by the impedance of air. It has been normalized by air and that is 5, 10, 15, and 20. And that is 5, 10, 15, and 20.

We could ask the question, are these reasonable numbers for the impedance of the ground? Since Hank is not here, we can do that. So let's look out here. The real part of the impedance is five and the imaginary part is

about five. We get the magnitude and say it is about five. Is that a reasonable number for the impedance of the ground?

If you have been listening to thermoacoustics all this time, what happens at the wall of the tube or the end of the tube when a gas parcel hits it? Velocity goes to zero. What is the impedance? Infinite.

This is not infinite, right? People at one point took a while to realize that the impedance of the ground really was finite, it was not infinite, and that is because it is porous, it is because it is a tube, so this is probably a reasonable number.

Forget about that line right there. I think there is a lot of structure in these data. I think it goes up and down and up and down. Most people do not believe that, that there is any structure in ground impedance.

[Slide]

This is a measurement of the geophone. This is the normal component of the particle velocity below the surface of the ground normalized by the acoustic pressure, so this has units of centimeters per second per microbar and goes up to about 100. It goes from 30 or 40 Hz up to about 250 Hz. Forget about the dashed dotted line but the solid line is the measurement, so you see some structure in this curve.

You primarily see two peaks.

The punch line is because the ground is layered in a weathering process. It is layered because there is a water table, it is layered because there is grass, the grass roots grow down into the ground and they stop somewhere, normally about a foot, and there are discontinuities, and those are the things that cause this structure.

[Slide]

This is the result of a probe microphone. That is not attenuation, it should say "magnitude." This goes from 40 Hz to about 2000 Hz. This is magnitude of the signal as a function of depth in 500-micron glass beads, which is like beach sand. This is the phase. What you observe in this measurement is that the attenuation at any one frequency is relatively homogeneous, but there is a tendency for the attenuation to increase. It is also on the order of about 5 dB/cm, so it is pretty big. Your sound is not going to go very far if your attenuation is 5 dB/cm.

This is the phase and you can look at the phase difference between two of these depths and come up with a measurement of phase speed and we see that the phase speed down at 50 Hz is less than 40 m/sec and it tends to increase with frequency.

[Slide]

This is taking the previous slide and constructing what the attenuation is and you can see here at about 40 Hz here is attenuation over the first 5 cm. This is the first 10, the first 15, and the first 20. This is in washed sand; that means that it comes from a sand quarry, it has a lot of clay in it, and they wash it, then they sell it to brick masons. It still has quite a bit of clay in it, but it is sand.

You can see that there is 2- to 3-dB variation in the data, it is not really very homogeneous as the glass beads were. The glass beads are significantly more repeatable than natural materials.

[Slide]

Let me say something about the properties of porous materials. We need to talk about the density of the solid and the density of the fluid. We can define the porosity to be the volume of the fluid divided by the bulk volume and, of course, the bulk density is one minus the porosity times the solid plus the porosity times the fluid density.

In the models that people develop for sound propagation in the ground they assume that all the voids are connected, that there are no isolated places.

[Slide]

Flow resistivity is an important parameter to be able to measure and if you have a cylindrical sample of length l and you apply a <u>delta</u> P across the sample and you measure the velocity of the fluid flow, you can define the flow resistance to be <u>delta</u> P over this velocity, the volume velocity per unit area, and that is like the resistance of a wire, we can get the resistivity by dividing by the length of the sample and there is an apparatus for doing this -- Leonard developed it and published it in JASA in 1946 and Rudnick made a modification to it.

[Slide]

Here is the jig. You can get this jig by going to the demo room and you take an old balance. This is the base of the balance, that is the pole, the knife edge is right here, and this is the mass bar. Usually you have a pan hanging on this side and a pan hanging on that side. You get the pans off.

You hang a cylindrical dish right here, a piston, if you wish, that you can attach to this side. On this side you can suspend weights. You have a sample holder you put your porous material in that mates to an annulus that you can put kerosene in. This piston now slips in this kerosene.

What you can now do is apply some mass here and you can pull this piston up and because it is sealed with the kerosene you can play air through here. There is a pointer down here and, as the beam moves across, you can measure how far the pointer moves and you can time the time that it takes the piston to go up to height h. If you get the time it takes the piston to go up to height h, you know the area, you know the volume of air that went through, you know this area and you know the length of the sample and you know the mass that you put on either side (you have to measure this thing), you know the force. On the other side is atmospheric pressure, so you can compute, then, <u>delta</u> P over the velocity.

In terms of all these parameters -- that is an expression for it -- mg <u>delta</u> t (the time) A_1 is an area, and there should be an A_0 here (squared) divided by the height it moves and the length of the sample. Right now I cannot remember which one is A_0.

If you take the sample holder off and you just had this open to the air and you try to balance this thing, what happens is it takes it a very long time to stop moving. When it starts to go down, this mass or volume that is in the kerosene displaces some fluid, so then the buoyant force increases, so it starts to go back up.

When it starts to come out, you have less volume in here, so it gets heavier and goes back down. What Rudnick did was put a little mass on top and you adjust the height of this mass to counteract this change in buoyant force. If you do not have this mass properly set, if you have it too high, it falls over, flops to one side. If you have it too low, it oscillates for minutes. If you put this mass on here it damps out in about two cycles (actually, about one-and-a-half cycles).

It is a very nice device, very simple, it works remarkably well.

[Slide]

Another property is the pore tortuosity. I have never made this measurement, but you fill the porous sample, you saturate it with a conducting fluid and you measure the electrical resistivity, then you do it with just the conducting fluid and you measure the electrical resistivity. They define the tortuosity to be this -- omega is the tortuosity of the sample.

[Slide]

In the next few slides what I am going to do is to develop two models of the ground. One is cylindrical tubes and the other is parallel slits. That is exactly what has been done by Larry and his graduate students in terms of looking at trying to develop viscous F-function models.

We want to develop two models of pores. One of them is cylindrical models of pores of radius R and the other one is parallel slits with 2A. We have seen this already, equations of state, continuity, and motion. We combine those to get the so-called most useless equation in the world.

[Slide]

There are some mistakes on yours. We take a solution (probably a useless solution, but maybe not) to find out that k, the wavenumber, omega/C, is this expression right here.

[Slide]

If we think about fluids with absorption, we usually write k as a complex quantity. We stick that in our solution and we get this, so we rewrite the wave equation with a complex C or a complex k, where k is in terms of this effective density and effective bulk modulus.

[Slide]

This is our goal, to write down expressions for the effective density and the effective modulus. If we can do that in terms of pore properties, and we know we can, because Anthony has done that for us already in his review, we can solve for the wavenumber.

If we have the wavenumber in this tube model, we can measure the phase velocity and the attenuation, which is what I had been measuring with the probe microphone. We are going to assume, to start, just as they do in thermoacoustics, that the tube walls are rigid, and when you squeeze the gas the tube walls do not move, but that is not what really happens in the ground. The ground really is elastic, so this is not complete enough to describe the ground.

[Slide]

The viscous penetration depth -- Anthony did that for us and I just put this here. At 100 Hz it is 200 microns. That is going to be larger than all of the pore sizes, or most of the pore sizes, we are going to work with. Normally, if you talk about spherical grains of soil, the pore size is about a third of the size of the grain, so grain sizes are from 50 microns to, probably, sand can be 300 microns.

[Slide]

This, historically, was done in terms of looking at ceiling tile by Zwicker and Kosten. Tijdeman did this, Pat Arnott did this, Mike Stinson did this, and Swift did this. The main thing that is assumed here in Zwicker and Kosten is that the pressure is a function of Z down this tube, and we talked about justifying that on the first day. We are going to start here and assume that that is the case.

[Slide]

I will just do a brief summary. For a cylindrical tube you write F=ma (Anthony did this). You rewrite this in cylindrical coordinates, you are supposed to say this is Bessel's equation and that is the solution, where l is defined to be this term here.

[Slide]

You now need to get the velocity across the radius of the tube, the average velocity, so you integrate it and we get that expression. Now s is the pore radius divided by the viscous penetration depth, the square root of two, and that has been called the shear wavenumber.

If you rewrite equation [30], then you can cast it in this form and you can see that we will write this effective density now as the original density plus this expression.

You can do the same the same thing for slits and you get a hyperbolic tangent, and s' here has a^2 in it, which is the half-width of the parallel slits. Biot did this, Attenborough did this, and their idea was that we are going to let these be the two limiting pore geometries of the ground.

We are going to say all soils must be somewhere in-between here and they scaled these two functions, and I will show that and introduce the pore shape factor so you can eliminate the radius or this semi-width and you end up with this adjustable parameter that is going to help you describe all porous materials. It does not work or it is not needed on the ground and I will show that a little further on.

[Slide]

This is from Allard and, in fact, my outline follows Allard very closely. This is the ratio of this effective density to the density of air -- that is the real part and that is the imaginary part. We are working down here all the time, so the low-frequency approximations for s are appropriate for those functions.

This is for slits. These two things look very similar. It leads you to believe that you can scale them and I will show the scaling in a second.

[Slide]

Anthony did this as well. We would like to get the heat-conduction effect, thermal conductivity, into the bulk modulus. I write it down for a circular tube and I write it down for a slit. We get very similar functions, except the Prandtl number is here, which is about three-quarters. The same approximations we used for the effective density work here. Hyperbolic tangents for parallel slits, that is what that function looks like versus s. Again, we are down in this regime down here. That is the hyperbolic tangent or the parallel-sided slit.

I am going to skip 10 slides in your books and go directly to the results.

[Slide]

What we were trying to do is, as you recall, come up with a bulk wavenumber in terms of the effective density, the complex density, and the complex bulk modulus, but we are going to do it in the low-frequency approximation. That is the only place we are really interested in.

[Slide]

If you do that, you can do all of this on a calculator. This is the result for the complex wavenumber. Gamma is the ratio of specific heats, a is this constant, aq^2 is the tortuosity. This S_p is the scaling parameter. For a cylindrical tube this is one, flow resisitivity, porosity, omega. What I do in the notes is show you how to get these measurable properties into the model of the ground.

When we go outdoors, we measure the flow resistivity and we measure the porosity, we do not measure the pore diameter, so we want to get the effects of pore diameter into a bulk media. That is the 10 pages that I skipped. I show you how to do that and it is not complicated. Again, this is the low-frequency approximation and it is good for all soils up to several kilohertz.

If we take the high-frequency limit of this expression, this thing gets small because of this term right here, this is constant. This is about one, the tortuosity squared is about one, and if this gets small enough that at high enough frequencies you can write this expression down. We see that this ratio of the sound speed to the sound speed in the bulk material goes as the square of the tortuosity. We call this an acoustic index of refraction. It does not agree with the measurement I showed you in terms of, conceptually, what is going on, but our notion of the tortuosity is it is the ratio of the sound speeds in air to the pores in the high-frequency limit.

Pat Arnott and I talked about what should the tortuosity be for spherical glass beads and came up with the notion that if a glass bead is a sphere, if the glass bead were not there, the sound would go straight through, but because the bead is here, it has to go around, so the tortuosity ought to be half the circumference divided by the diameter, which is $\pi/2$, or 1.6. That is a generally accepted number for measurements of tortuosity.

[Slide]

The problem with k_b is it is in this square root sign with this expression. What I can measure is the attenuation in phase speed in the ground and I would like to back out the properties of the ground. I would like to determine what is the flow resistivity of the ground, what is the porosity.

If you just squared this, then you can separate the real and imaginary parts of this expression. You can equate, then, the real part of k_b^2 with frequency squared and tortuosity. The imaginary part of this expression has <u>omega</u>^{2} in it and <u>sigma</u> <u>omega</u>.

The only unknowns in this expression are <u>sigma</u> times <u>omega</u> and the unknown here is q^2. I can use my probe microphone measurements to measure these properties -- only two of them, I cannot get all three of them.

If we look at the impedance, the normalized impedance is one over the impedance of air times the frequency divided by the square root of these two terms, the effective density and the bulk wavenumber. In the low-frequency approximation of this expression we find that in the low frequency the real part of the impedance and the imaginary part of the impedance are equal and they go as one over the squared <u>omega</u> -- <u>sigma</u> over <u>omega</u> square root, one over this frequency.

In the high frequency we see that they are proportional to a constant, the square root of the tortuosity squared divided by the porosity squared.

We set off not understanding this very clearly -- this result right here -- trying to measure all the properties of porous materials. We were going to use reflection measurements, and I will show that, to determine the tortuosity, the flow resistivity, the porosity, and even the pore shape factor, before we knew how unimportant the pore shape was.

This tells us that at least in impedance measurements you cannot do it. These are not independent parameters. In the high-frequency limit you get a ratio of these two parameters. In the low-frequency limit you get a ratio of these two parameters. They are not independent, that is what I am saying here.

[Slide]

This is the calculation of that previous expression for the impedance. At the low frequency we see that they are very nearly the same. At the high frequency (this is not quite high enough), this one is tending to zero and this is going to tend to a constant. It becomes purely real.

The data tend to have this correct frequency dependance.

DR. MIGLIORI: How can you fit a horizontal line through that?

DR. SABATIER: I did not say it fit the data or, if I did, I did not mean to. I still contend there is a lot of structure in these data. This assumes that the ground is semi-infinite material of cylindrical tubes.

[Slide]

I mentioned this probe microphone. Here is a closeup of it. You buy this element at Radio Shack for a dollar. You put it in a brass tube and you have to design a nose cone to go on it. The only reason you put a nose cone on it is to protect the element when you push it in the soil. It is invasive, you have to be very careful when you insert this. In glass beads and sand it works perfectly, because they seal. When you do it in soil you have to drill a hole and you can have leaks and it does not work worth a hoot.

Here is a photo of it. You use a camera tripod and this acts as a jig to drill holes in soils and also to put the probe in to support it, and there is the reference microphone.

I am going to skip this, you have seen these kinds of data.

[Slide]

Here is the analysis of the data. We measure the probe response at two different depths due to a source in the air as a function of frequency and we compute a transfer function. From that transfer function we can get the phase, we can calculate the phase of the transfer function. That is equal to the real part of the wavenumber times <u>delta</u> d.

We calculate the magnitude of the signal dB and we can relate that to the imaginary part of the wavenumber. We can solve, then, for the real and imaginary parts of the wavenumber from our measurements; we can construct k_r and k_i.

This is a plot of the imaginary part of k_b^2, these are the data, and that is a fit.

[Slide]

This should be linear with frequency. We determine the slope and with the slope we can come up with the imaginary part of k_b^2 -- right here -- we can determine the product of sigma omega.

[Slide]

This is a fit, then, to the real part of k_b^2 and that is the one that should go as omega^2 despite what is in my notes, so we can fit this and we can determine the tortuosity.

[Slide]

These are measurements for five sizes of spherical glass beads that range from 500 microns down here to, I think, 60 microns up here in diameter. The important thing is that this is the DC flow resistivity that we measured with Leonard's apparatus, and this is the result of fitting sigma omega to the data.

This is the sigma omega that we calculate with the model and the probe microphone. This is the DC flow resistivity. If you take the slope of this line you get 1/.4, which should be the porosity. The slope of this line is 2.5. One over that is the porosity. The porosity of randomly packed glass beads is 38%, so this technique arrives at a reasonable number for what the porosity of the sample is or, if you know the porosity, it arrives at a reasonable number of what the bulk flow resistivity is. It is a way to measure the flow resistivity of the ground, but you have to put the probe microphone in the ground and that can be a problem.

[Slide]

We can take advantage of another measurement (Ken talked about this). This is a sound source, and this is a microphone. This is a Lloyd's mirror experiment. We broadcast sound from here. A ray goes there. A ray also goes up here and down there (this is the reflected ray). There is also a surface wave. It is all in the model that Ken put up for this.

Here is the pressure at that microphone. It is $e^{ik\ 0}$ times r_1 divided by r_1 plus the reflection coefficient times $e^{ik\ 0}/r_2$. This is done with spherical waves, so this is the spherical wave reflection coefficient, which is written in terms of a plane wave reflection coefficient and this correction factor, which is a function of impedance and frequency and angle.

Z is the function of these three parameters. What we do is we broadcast swept tones, swept sines or noise, and we measure this response. We usually put another microphone right here and take the transfer function of these two to take out the source.

Once we do that, we can then fit this model to the data and tweak these parameters to see if we can do Lee squares inversion on the data and come up with those numbers.

[Slide]

Carl Fredrickson did this. Here he has done the calculation of the frequency response from 10 or 20 Hz up to 2 kHz for the sound level difference for two of these microphones, two vertically separated microphones. He did it for cylinders, slits, triangles, and it is not dependent on pore shape. That is why we believe we do need pore shape.

He now does the calculation for bulk flow resistivity and he lets the flow resistivity range from 69,000 to 80,000, and he shows that this is the kind of effect you are going to get, so we are not going to be able to get the flow resistivity extremely accurate, but if the flow resistivity is much larger than this, it is a dramatic effect; this thing can change easily by 20 dB, over two orders of magnitude in flow resistivity.

Here are some measurements of this level difference measurement, and that is the calculation of that curve, the Lee squares fit. From these data you can get the ratio of the pore parameters that I wrote for the impedance. Carl did this and he compared the probe microphone measurements of the properties of glass beads to the level difference, to the reflection technique, and it did not work very well, and he did it for sand.

The reason that it does not work (he did not learn this until after the fact and it has not been done again) is that when he put his sand in, he just threw the sand in with a shovel -- actually, a front-end loader, I think -- and then he took a skrete and he scraped the surface until it was smooth; he did not make the material homogeneous, he just put it in like it was.

Then he made probe measurements down to a depth of 20 cm and then he made level-difference measurements. The level-difference measurement does not sample nearly as deep, it gets mostly the surface properties of the porous material, so when you go with a probe over deep depths, if the material is not homogeneous you cannot get the same measurement as the reflection -- at least I think that is the estimation.

Now we are going to take a 10-minute break.

DR. SABATIER: The second part of this has been that the ground was rigid -- maybe I did not emphasize that enough -- and the granular materials could not move, only the gas could be moving in and out of the pore space.

[Slide]

When we put a geophone in the ground, the geophone responds to motion of the matrix of the soil, so we have to do something different when we find mines. The laser is measuring, I think, the velocity of the soil particles, the particle velocity of the soil itself on the surface of the ground.

The way we do that is to go through wave equations developed by Biot. I want to point out a couple of things in these wave equations; these are coupled wave equations. H, C, and M are elastic constants that can be defined (that Biot defined) in terms of the complex compressibility -- he did not do that but Attenborough put in for air the complex bulk modulus in terms of M and C. H just looks like lambda plus 2nu in some limits of poroelastic materials.

Here is the F function written in the F form that other people have talked about. This is the flow resistivity.

First of all, e and squiggle, e and Z, are displacements -- actually, dilations -- of the solid in the gas of this poroelastic material. This term here has been referred to as an inertial drag term. It is equal to the tortuosity, the fluid density divided by the porosity, so if this term would go to one -- well, let me rephrase it.

Because the measure of the tortuosity of the pore path, as the air tries to move back and forth, curved pores, as you accelerate gas you accelerate the wall, this is a measure of that effect -- it is clearly a very small effect.

[Slide]

The stress-strain relationships: The only thing I want to point out is that there is a connection between the pore fluid pressure and the dilation of the solid and the dilation of the gas. On this slide I want to point out if we take plane wave solutions to these equations and stick these back into the two differential equations, we get these characteristic equations in which we know the determinant has to be zero and the solution is in terms of l^2, or it is quadratic and l^2, which means there are two roots, two wave types that result from these two equations.

It kind of makes sense. You have two phases, but they are coupled. In one of these solutions, say we call it Biot type I, the solid and fluid are moving together as a wave. In the type II wave the solid and the fluid are moving together but with a different wavenumber, so there are two wavenumbers and these two wavenumbers have different dispersion characteristics. Their characteristics are defined in terms of a type I and a type II wave.

If we solve either of these two equations for the ratio of A/B -- so what are A and B? A is the displacement, if you wish, of the solid part of the matrix, B is the displacement of the fluid part of the matrix. If we wrote these as particle velocities, then it could be velocity amplitudes. It does not really matter, it is just time derivatives.

Here, then, we can solve for what the ratio of the amplitudes is of the two wave types and there is an l here that is a wavenumber, so there are two wave types, there are two amplitudes.

[Slide]

On this slide we plot the phase speed and the attenuation for the type I wave and we put in properties that are appropriate for saturated sand, air-saturated sand. This is the type I wave and we find that we get a phase velocity that is constant with frequency in the range we are working over and we get an attenuation that increases as F^2; however, it is very small.

This is the ratio of the solid displacement to the fluid displacement for the type I wave and we see it is about one, it is on the order of one. That means that the fluid and the solid are moving with the same amplitudes for this wave type.

If we go to the type II wave and repeat the calculations, this is just the phase speed and the attenuation that comes from the tube model. Here is the ratio of solid-to-fluid displacement and now it is very, very small for the type II wave. What this says is for the type II wave the matrix does not move very much compared to the gas. In the other case they are moving together.

[Slide]

This is an experiment that was set up by Craig Hickey to look in the laboratory -- I did all this stuff outdoors and he was not happy with that, so he wanted to look at what happened in a tank if you could try to do a controlled experiment.

DR. ATCHLEY: Can you go back to the last one? Can you go back and explain the difference between them, again?

DR. SABATIER: Yes, I can. I did not do a very good job, now that you ask me.

[Previous slide]

The type I wave, when you solve this equation, you get two roots, you get two wavenumbers. One of the roots corresponds to in-phase motion of the solid and the gas. This is the displacement of the solid for the type I wave. This is the displacement of the fluid for the type I wave, so this wave propagates and squeezes the gas when it propagates. They go together, they are coupled.

DR. GILBERT: This is like forward-going and backward-going. You do not know what you are going to get until you do something physical.

DR. SABATIER: That is just solutions to a wave equation, no boundary conditions. It is an infinite material. If you do it for l2. there are two of these, so now there is another wave that goes into the solid with a different wavenumber and coupled to it is a wave that goes into a gas with the same wavenumber, with the same phase speed.

The l1 and l2, these are the wavenumbers, so l1 has little attenuation and a constant phase speed, no dispersion, and l2 is the tube wave that we have been talking about. It has a lot of dispersion, its velocity changes with frequency, and its attenuation increases with frequency.

A lot of people believe, or at least say, that the type I wave moves only in the solid and the type II wave moves only in the gas, and that is wrong, and I am going to show that in a measurement that that is wrong.

[Slide]

Let me go to measurements that I think are going to show some of this. This is a tank experiment set up in the lab. This is meter by meter, a cylindrical tank. In this tank Craig buried vertical component geophones down here and he buried them in 10-cm spacings on this side, and on this side he offset them by 5 and buried them in 10-cm spacings, and I guess they are 30 cm from the wall.

Then he put in microphones all the way down to the bottom of the tank, also in 10-cm spacing. He put in horizontal component geophones, so these vertical component geophones respond to motion this way, the horizontals respond to motion this way.

He has horizontal, vertical, and geophones and microphones. Then he has two sources of excitation. One is a loudspeaker suspended above the surface. The other one is a mechanical shaker that he can put on the surface here or on the surface there -- or anywhere he wants, actually.

[Slide]

No one has been able to separate out a pulse for these two wave types unless you go to low-temperature physics, and I do not know if they did it there, but in water-saturated porous materials and air-saturated sediment you cannot directly separate out the wave speeds.

Plona did some tricks, but he did not separate them out going down through a porous material; he had to take advantage of a critical angle [?]. He is the first person to make a measurement of this slow wave and report it in the literature in a poroelastic material.

What I will say is that everybody who did architectural acoustics and everybody who does thermoacoustics have been measuring slow waves as long as they have been making measurements, at least in the gas, and thermoacoustics people do not worry about the wall of the tube. In architectural acoustics people have only recently begun to worry about the walls of their architectural materials.

[Slide]

Here is the loudspeaker source and he sends a 3-cycle tone burst at these three frequencies. He measures the time of arrival versus the depth in the tank. These are shallow depths and at shallow depths, at the first 10 cm, all he has is the probe microphone -- I failed to mention that. Let me back up.

[return to Slide]

He has a probe microphone that he can look every centimeter with, so with his probe he can look every centimeter down this tank.

MS. SWEARINGEN: Did you mention what material was poured into this tank after all these sensors were put in?

DR. SABATIER: These data are with one of the sands that we used in the previous experiment, so it is sand that has, probably, grain sizes of a few hundred microns in diameter. It is very clean sand.

[return to Slide]

This is the arrival time at the sensors as a function of depth. These are three frequencies where he broadcasts a tone burst. He plots the arrival time versus depth, so one over this is the wave speed. These are the slopes of the data, and it is 122, 134, and 149 m/sec.

This is the kind of dispersion that you would expect from the slow-wave calculation. From the rigid porous material calculation this is what you expect.

[Slide]

He now uses the shaker source, not the loudspeaker, and with the shaker source at these three frequencies, using microphones and the vertical geophones, he is now looking from 5 cm deep into the tank and he looks at this arrival time (that is probably milliseconds, again) and he has, on this curve, a microphone and a geophone at every place -- the same thing on this one and the same thing on that one, and these are the same three frequencies and these are the velocities. They are either constant or there is negative dispersion (I would like to think they are constant).

You do not see dispersion and the velocity is twice as high, and it is twice as high all the way to the bottom, all the way to the top. This is a mechanical shaker shaking the source instead of a loudspeaker.

[Slide]

This, now, is the loudspeaker source. I showed it to 10 cm earlier. Now this is the loudspeaker source all the way to 50 cm and he gets these numbers at 1 kHz. He gets these numbers back at low frequency, 143 m/sec. Then there is a transition region to the deep sensors where you get 240 m/sec.

You see that on the probe microphone, which samples every centimeter down to about 20, and you see that on the buried microphones all the way out to 30 or 40, and you see it on the geophone at 5, all the way down here but, importantly, right here.

[Slide]

He computes the sound absorption, the attenuation, and I hope yours says magnitude of the signal -- I transferred these over the net and somehow a lot of things got lost. Two kilohertz -- this is the attenuation now, this is the loudspeaker, and he is using the probe microphone and he gets this straight line, and that straight line, 3.8 dB per centimeter and 1.3 dB per centimeter.

If you measure the flow of resistivity and the porosity and make some assumptions about the tortuosity -- take a guess, we do not measure it other than acoustically

-- these are the numbers you get for the slow wave.

MR. APOSTOLOU: There are two different slopes, but how can you tell if one of them is V1 and the other one is V2?

DR. SABATIER: We do not know which one is which. Oh, how do we know? Because we showed that in the shallow depth there was frequency dependence and V2 has a lot of dispersion and V1 does not.

The important thing here is that when you use a loudspeaker, the loudspeaker is impedance-matched well to the air in the pores, so it drives the gas in and out, but the microphone and the geophone are both measuring the same thing.

I will say that a little differently. The microphone and geophone are both giving us the same phase speed. One, I think, is measuring matrix velocity and the other one, I think, is measuring fluid pressure.

PARTICIPANT: [Inaudible]

DR. SABATIER: Time in milliseconds, arrival time, so this is like inverse sound speed.

This is important in that if you use a shaker you do not get this. I do not know whether he did not take the data, but I do not have them.

We have done this on two different other kinds of soil types and if you have a much higher flow-resistivity soil, this break point moves down deeper and deeper. All I am trying to tell you is if you use a loudspeaker, slow wave is the predominant energy and that energy decays away and you are left with another slope, which is the fast wave, or the type I wave.

In big enough pores the slow wave can get faster than the fast wave, so that is not a good choice; the type II wave can become faster than the type I wave if your pores are large enough. In granular materials the sound speeds is nominally 200 m/sec. If you get big enough pores you can get above that.

Let me back up a second. I think this idea is important in terms of how we are finding mines. If you use a shaker to excite the ground, you do something different than if you use a loudspeaker. In one case you can excite this type II wave, in the other case you cannot excite the type II wave.

MR. APOSTOLOU: You think you are getting good success with all kinds of soil because you couple through the gas, is that the reason?

DR. SABATIER: No, what I want to say is that the type II wave can see this target. It is very sensitive to changes in porosity, the tube diameter, particularly if you close off the tubes. That is what I want to say. There was no notion about mine detection when we did these measurements.

The motivation for these measurements was we knew that if we took our probe microphone and pushed it down in a soil and measured the attenuation very carefully as a function of depth, we got nice uniform attenuation. Then at some depth we did not get uniform attenuation. The signal would go up, it would go down, it looked like noise, you had a sine wave. You turned the loudspeaker off and there was still plenty of energy.

When you do this outdoors in the ground or when you do it in this tank, actually (you can see this effect in this tank, I think, as well), the type I wave has a lot less attenuation. The slow wave just dies out, it does not know what is below it unless it is very, very close. With the fast wave there is much less attenuation and it sees boundaries that are below the surface of the ground. That is what motivated us to try to do this experiment.

When you take the probe microphone and you start measuring the pressure as a function of depth, it decreases, decreases, decreases, and decreases differently, and that is what prompted us to do this. We are doing this again now and we are going to do this in five sizes of glass beads and eventually we going to put land mines in here. Glass beads have nice properties because they are all the same as you change the size, except for the flow resistivity.

[Slide]

This is the notion of what the ground looks like. We want to apply this model to the ground. This is the air-soil boundary. I just assumed that the ground was layered, just a single layer, and these are the wave types that are allowed for a plane wave source, an incident wave and a reflected wave.

This is the type I wave. This is a porous material and this is the displacement or the amplitude of the motion for the matrix -- that is for the fluid. It hits this boundary and reflects. I assumed that this was non-porous, so it transmits only as a type I wave. It could, of course, mode-convert here to a type II wave upon reflection.

Here is the type II wave coming down. This is the amplitude or displacement of the matrix for the type II wave. This is the displacement, B_2, for the displacement of the type II wave for the fluid. It can hit here and reflect. It can mode-convert to a type I wave, but it cannot go through the boundary, because I am assuming this is non-porous. This is a shear wave, and I did not bother to show those equations.

MR. SPARROW: I am curious....local reacting assumption at this point?

DR. SABATIER: Yes, in fact, I never use the local reacting assumption. The local reacting assumption, and you might correct me, assumes that the sound speed tone is to the normal. What does the speed of sound have to be in the ground for that happen?

MR. SPARROW: It is slow.

DR. SABATIER: Very slow, zero maybe. But Pierce talks about local and extended reaction and it is a little different idea.

So we need some boundary conditions and I am not a mathematician. There is a series of 13 papers about the boundary conditions between poroelastic materials. They are in the Bulletin of Seismological -- something -- which I cannot remember right now, by Doritzovic, I think, and somebody named Scaflak, and I refer you to them.

The top set is just general boundary conditions for a porous material, continuity of, the normal component of stress, matrix of velocity, parallel components, tangential components of stress, velocity, normal components of pressure, and fluid velocity.

Over a pore-solid interface, I list them there, and over a fluid over a pore-solid, I list them there, and I used those boundary conditions in the calculations I am going to show you (I took them from the literature).

[Slide]

There are 10 amplitudes in the previous sketch. One of the amplitudes is known; that is the incident. We can apply the nine boundary conditions that we have, the four for the fluid over the pore-solid, the others for the pore-solid over the substrate, non-porous substrate, do lots of algebra, and give that to a matrix-equation solver and solve for the amplitudes.

Once we have the amplitudes solved, then what we have to say is suppose we want to calculate the normal component of the particle velocity, we have to come down here and sum up the normal components of all of these plane waves and add them up. We do the a's and we can get the matrix velocity.

We could do the same thing for the fluid. We can do the same thing for the incident reflected wave. We can construct things like the surface impedance and this should not be a 1 and a 2, this should just be pressure at the surface is ik_0 for air, l_0 for the wavenumber in air times the incident reflected amplitude, the sum of those two. It is just the sum of the pressure above the boundary.

We can do the same thing for the surface. Scratch out these 1's and 2's. These are incident reflected amplitudes, that is the incident angle of the wave, and the other thing here is this kl_0 for air -- l is for wavenumber in air, omega/C_0, k is the complex compressibility of air, which is the compressibility of air, it is not complex.

Once we have P and V, then we can construct the impedance, the surface impedance. If we want to measure what happens in the porous material, we can sum the pressure associated with the down-going and the up-going waves. I already said if we want to find the particle velocity as a function of depth, if we want the normal component, we can sum the normal component of the up-going and down-going type I and type II plus shear waves.

That is what I did and these are the results.

[Slide]

Where did I get all the numbers? We can measure all the pore properties. We can measure densities. To get Biot's elastic constants we have to get the matrix compressibility, we have to get the granular compressibility, and we have to get the shear moduli.

What we do is do something called a shallow- refraction survey. You hit the ground with sledge hammer and you have a string of geophones and you measure the arrival time of the wave speeds. If you string your geophones far enough apart, the wave speeds increase in the ground and you will see deep-layer wave speeds -- this is what seismologists do to find oil.

We do that experiment and that lets us find all the velocities in these two layers. With those velocities we construct Biot's elastic constants. What I will tell you is that the grain modulus does not matter. It does not make any difference what the grain is made of, within reasonable numbers. So these are measurements, and these are calculations.

I should point out that Pat Arnott made the same calculations using an elastic theory and gets the same results, so you do not need, on this particular soil, Biot theory to get these results, but on Craig Hickey's soil (and that experiment still needs to be done) it suggests you need Biot's theory or it suggests Craig Hickey is wrong.

[Slide]

This is another material where the measured is the solid line with lots of peaks and valleys and there are no adjustable parameters. You put in the layer depth, you put in the measured moduli via the speed sounds and you get these kinds of curves.

[Slide]

This is the interesting stuff. This is a calculation of the surface impedance from this layered model done at two different angles of incidence, 80° and 68°, and you get some funny structure in the impedance curves.

[Slide]

If you take the layer velocity and you bump it up by a factor of 10, the substrate velocity, the curve does this at 0° and 80°. At 0° the solid line shows one peak here, another one there, another one there -- this is on a log axis, log frequency, and it goes from 50 to 1000 Hz.

When you go to 80° you get a twin structure, you get two peaks; they come in pairs. These are resonances that are due to compressional waves and shear waves going down into the porous material, but there is a smooth trend in the impedance to decrease with frequency as well as in the imaginary part. I showed both of these to Henry and he said to make them all go away.

People have measured them and some people think they see them in some cases. In snow they are clearly visible but no one -- Giles Dagle, maybe, published a paper that says he measures them, not this, but that.

[Slide]

Hank's data are right there. It seems to me we need to find out where these data were taken, but I have a feeling that the layer depths are going to be -- we could almost fake it and begin to show that we are going to get structure in these curves.

I think this is all I want to do here and I want to switch and do something else.

[Slide]

What we have been doing for the last 18 months with this technique for finding things buried in the ground is to try to find out how well we can make it work, or where won't it work, under what conditions.

One of the questions we were asked is how far up the road can you find land mines or things that are buried in the ground. I mentioned that we tried to get the vibrometer away from the loudspeaker. Here is an attempt to put it in an isolation box (Dan Costley did this). In this box the vibrometer sits on a set of springs and this is some constrained-layer damping material in an attempt to try to quieten the vibrations, reduce the vibrations. A lot more work needs to be done.

In the next set of slides I am going to show you what happens. The mine is over here and we are going to back this forklift up to get farther and farther away. We are going to keep the loudspeaker in the same place to keep the sound pressure constant.

What is going to happen is the angle of incidence of the laser beam is going to come down and we are going to be measuring the cosine of the normal component, but we are going to be getting part of the radial component of the velocity.

[Slide]

We also put these up on this device up here and that makes the coupling worse. We have had this thing operating as high as 7 m, looking 7 m in front of us. When we know there is a mine there, we find every one of them and we go to the calibration lengths. It works very well.

[Slide]

Here is the particle velocity measured at one point on the target -- this is velocity as a function of frequency, 50 to 500 Hz. This is the angle of incidence from the normal. This is about 2 m, and this is about 12 m. As you move away, the velocity gets smaller, about like the cosine of the angle.

There is a problem, and that is if you look at the off-target velocity, this is the one at 46°, as you move away, it does not change. This is one of our difficulties and we are working to improve this, but we are probably not going to. As the vibrometer looks farther and farther away, not only does the velocity come down, the noise floor of the instrument comes down, because there is significantly less light coming back into it.

As less light comes into the optics and eventually through a phase-locked loop detector, the noise floor comes up. These things will measure 50 nm, but that is under optimal light conditions and we are not under optimal light conditions. The light coming back is always at a very low level in this instrument.

We did a blind test in a forward-looking scenario and we decided to do, at about this angle, a 6.5-m standoff in front of this vehicle. We did it on the same lanes where we did the first blind test, where we found 95% of the mines. We found 70% of the mines and we had no false alarms.

I think it is correct in the sense that the noise floor is coming up and we cannot see mines when the velocity is close to them. As this thing comes up, you cannot see these mines, but they do not look like mines. It is random velocity variations over the surface and they do not appear as big round red circles, so we need help here and we are working with some people to try to improve this.

[Slide]

This is what an image looks like as you start to look off; this is at 8 m, and you begin to get some distortion in what it looks like.

[Slide]

People do all kinds of things to us now. Somebody decided that this is how an anti-tank mine is buried, and this is how anti-personnel mines are buried to protect the anti-tank mine.

Maybe you find this thing with your sensor, but when you walk up to it, you step on one of these -- kind of a nasty trick. These are real, these are live. I had the opportunity to glue some accelerometers on these, there are two here and there is one on the bottom, gljued on top of it; I was just trying to see what the acceleration looks like.

[Slide]

That is an image of what it looks like. This is up to 300 Hz. These are the three anti-personnel mines. This is the anti-tank mine. Normally, these things show up at around 150 Hz. When you go up to these high frequencies you start to get this kind of distortion. Maybe with this characteristic pattern it could help to find mines that were deployed this way.

[Slide]

Here are some intriguing things that might explain a little bit about why I think the slow wave is what is seeing this mine, or at least shallow mines. These are two mines that were buried -- all the mines that were buried have been buried at random depths. I requested that they be buried at 1-inch increments, take this mine and bury it from 1 to 6 inches in increments of an inch, and do the same thing for that. This is a foot in diameter. That is about 10 inches in diameter. This is 4 inches, and it stands up about 3 inches on top of this mine.

[Slide]

This is the particle velocity measured with the vibrometer on top of the mine, right in the center of this VS, and this is its depth, from 1 to 6 inches, and this is the background. What you seem to see is, first of all, signals that decrease in amplitude with frequency -- or with depth -- and signals that decrease in amplitude with frequency, so you see attenuation due to depth and you see attenuation due to increased frequency.

You also see attenuation here, but no attenuation there at the deeper depths. The same thing occurs on this target. This target is not nearly as visible -- I suspect because it is automatically 3 inches deeper. If we look in this low-frequency band, we do not think we can see that 3- or 4-inch circle, but we can see the big body of the mine, which is a lot deeper, so we start off with the smaller signal.

Maybe there is a decrease in amplitude and maybe there is not. The peaks do not move around, I will say that. The deeper the target -- this happens to be, I think, the deepest target -- these are two others. this one is a foot across, that one is about 10 inches across. These are anti-tank mines.

[Slide]

Here we see a similar phenomenon. These two are attenuated, but these are not, these are more random -- two are obviously not enough. Maybe it is just as bad here. Here is the background, the dashed line. The green is the 1 inch, the blue is down here, but the 5 is almost as high as the other one. If we have a lot of signal, we seem to see this phenomena a lot, where there appears to be two kinds of attenuation. When we do not have a lot of signal, we do not tend to see that.

[Slide]

These are some intriguing things about shape. If we stand these three objects buried in sand about an inch deep and we do a very high resolution, we scan them with a pixel size every millimeter, everything else has been scanning about 10 cm, when we stop every millimeter and collect data, we get these kinds of pictures.

[Slide]

The round disk is round, but the amplitude is positive, or higher than the background -- you actually see another ring around here. For the squares and the triangles it is smaller over the target, higher around. This is a single frequency, the velocity over the surface. This is 8 inches by 8 inches, a 4-inch target.

[Slide]

If we attempt to do a nearfield calculation of the pressure and say we are going to calculate the pressure in a fluid for a target of that size and we do this equation numerically --

[Slide]

-- for the long wavelength case, if the wavelength is three times the target size, the diameter to the side is one and the depth is a 10th of the diameter, we get a square, a circle, and a triangle, and we see dips over the center, and we can see no dip here, so we can find frequencies where we see the behavior that I showed you -- these have holes in the middle, and these have bumps in the middle and it is bigger here.

[Slide]

In the short wavelength case, where the wavelength is a third of the target size, the diameter of the target, the depth is a 10th, we get this kind of phenomenon.

[Slide]

I showed that one, and I would like to stop and take questions. Thank you.

DR. MIGLIORI: What do they do when they find a real one?

DR. SABATIER: You put a mark on the ground.

DR. MIGLIORI: So you are driving along in this truck and you say, "There's a mine," and then you decide to miss it if you continue to --

DR. SABATIER: You have a vehicle that can drive over it, it does not bother it. It has enough surface area that it will not set it off.

DR. MIGLIORI: Even the anti-tank ones?

DR. SABATIER: It will not set off the anti-tank ones. The problem is, it will set off the anti-personnels, but they will not bother it. There are a lot of overpass vehicles now that are currently already at work.

DR. WILEN: When we saw the animation, it was obvious that the signal over the mine was out of phase with the ground and we did not get to see an animation for the rocks. Is that the same case with the rock?

DR. SABATIER: It does not do that. You do not see this phenomenon on every target. You see it on lots of them, though. Every mine does not behave that way, but lots of mines do.

DR. ARNOTT: On that calculation you did for that, could you say a little bit more about that, the one you had two slides back?

[Slide]

DR. SABATIER: Craig typed that into Mathematica and he just let it be -- it has been done two ways. The first one was to just let it radiate, the piston radiate. Subsequently, we have been working with doing the floor. He has a circular disk, he lets a plane wave come in and reflect off the surface and he sums them up.

What we know is that as this gets deeper, this effect goes away very, very fast. It does not have to get but about half the diameter deep and it looks as if everything is round, but up when it is very close, you actually see the shape, so it is just done on Mathematica.

The integration routines you do not do on Mathematica. You go to some library somewhere to see if you can find a Mathematica that will do the integrals over all these shapes very fast, so we did not sum them up point by point. We used someone's algorithm developed for integrating shape.

DR. WAXLER: I remember a few years ago hearing about some people who were studying sand and they noticed that it seemed to come in clumps and the conjecture was that the stress holding the sand up occurred in these two-dimensional sheets. I wondered if you see anything like that -- I guess in type I wave scattering.

DR. SABATIER: I think there is definitely something referred to as stress chains through materials, where the materials are connected together. You move the material below it and it will stay in place.

At the frequencies we work at, we do not have, I think, the resolution to resolve that, would be my guess.

MR. APOSTOLOU: When you say "slow," you mean type I?

DR. SABATIER: I mean type II, I am sorry.

MR. APOSTOLOU: Is 6 inches the maximum depth they bury these things?

DR. SABATIER: That is too deep, more like 1, 2, and 3 inches is more typical. How deep people bury mines depends on how much time they have. In some places you cannot bury them at all, they just cannot get them through the ground.

DR. GILBERT: You described to me at one time the original test that you did out in Hank's back yard with a sledge hammer. You hit the ground and then, a long time later, you would see this pulse come in that was more like 50 m/sec, but weren't you getting a seismic wave really quick and isn't that a measurement of the slow and fast waves?

DR. SABATIER: Kind of yes and no. Henry reported that. You fire a pistol and if you have a geophone over there, you see a wave that comes in at a very high speed with a very low amplitude and then you see a wave come in with a high amplitude that corresponds to the speed of sound in air.

Now, the wave that you receive at that sensor is the super position of the type I and the type II waves from air. The one that you get from the pistol hitting the ground, from the pistol signal going through your body into the ground -- and it goes deep into the ground, where there rocks -- it travels at 3 km per hour, very fast, and then comes up to your sensor, that is the only fast wave, seismic wave -- actually, it is probably motion of the gas associated with it.

The main thing is it is different excitation mechanisms is what I am trying to say. In one case you are preferentially exciting the slow wave. In the other case, you do not excite it. It is not that you are not exciting the fluid and the gas; you always excite the fluid and the gas.

DR. GILBERT: At a long enough distance you ought to be able to separate the pulses if it does not decay too fast.

DR. SABATIER: It decays extremely fast. In the soil behind my house, when we cut the grass and kill it and spray it, and let all the organic matter in the grass decay away, it is soil that has about 40-micron grain size, the attenuation at 100 Hz on a good dry day is 10 to 15 dB/cm.

On a wet day, when it is really wet, it can go up to 40 or 50, but you can still find mines with acoustic seismic coupling. We have taken that when it was pouring down rain, it does not make any difference. We have taken it after long rains. That is probably just the fast waves. Both of these waves can see the mine. That is the point I am trying to make, but it is shallow, and if you have a low enough floor sensitivity, the slow wave really sees them, but we cannot predict the frequency of these curves, we cannot do that.

We are working on it, but we have to solve the scattering problem of a non-porous elastic material -- a cylinder in a poroelastic material. It has been done in underwater acoustics in a way that I am not able to understand -- by Ray Lim and Steve Cargill, and we have invited them to try to help us work on this problem.

MR. GLADDEN: You say right now you are at about two minutes for a scan on a good day --

DR. SABATIER: When the thing is working right.

MR. GLADDEN: Have you thought about any ways that you might speed that up? I am sure that is a concern.

DR. SABATIER: Sure, we have. I forgot to show a couple of slides, since you raised that.

[Slide]

We have taken a mirror and we put a triangle wave voltage on one of the mirrors on this LDV, so it sweeps across. This is a very low sweep rate, 16 seconds. As the beam sweeps across here, this drives the beam, this is the velocity output of the vibrometer and the mine is right here in the middle.

Two things happen, and we are broadcasting a 180-Hz tone, single frequency, so we get some big spikes, but there is a general increase in the signal over the mine and then it goes back down again. When the voltage tells the galvanometer to go back the other way, it rings on both ends when it stops and when it goes this way -- when it stops, I guess.

This is the spectrogram. There is a postdoc named Vincent Vallo, from France, who has been doing this. Here he is doing a spectrogram, it is a timed frequency analysis, because you have to take a Fourier transform with time. Horizontal energy correlates to what we are broadcasting. Vertical energy seems to correlate with the speckle -- or at least we think this is a phenomenon called speckle.

Here is some other kind of distribution and it cleans this picture up a little bit. We have been able to show this work on both the roads at A.P. Hill. We are broadcasting now at four tones and here you can see the four frequencies. This is the spectrogram. This is one of those big pulses that has nothing to do with the mine, it is continuous in frequency. Notice this is 4 seconds and this is a meter -- we are scanning a lot faster now.

[Slide]

This is something called the short-time coherence and this is the spectrogram time of the coherence, so these are signal-processing ways of finding mines. This corresponds to 1.4 km/hr, this sweep rate.

What we are going to do is we are working with a company that is going to put 16 700-nm Doppler beams through one Bragg cell and we will have 16 beams coming out. The reason we are going to do one Bragg cell is you save a lot of money. The Bragg cell is $5000, the most expensive part.

That is going to let us have 16 parallel beams and we are going to be able to go at 1.4 km/hr, that is what we think, that is what we are going to try to do. This shows that we can find mines where the beam moves. People who have vibrometers say this does not work, you cannot do this. It works, okay? It does not work very well, but it works.

Now we still have to deal with the problem of the platform is not moving, so that is going to be another problem and we are going to have to deal with that. We can find mines when the forklift is running; we did that accidentally. Dan Costley is in the room, Dan has worked with military technologies. They are a contractor in our building and we subcontract to them. Dan wrote the specs on a ruby laser wholefield interferometer.

This thing sends out a collimated light beam that is a meter by a meter and it is pulsed and has a CCD camera that catches the reflected speckle pattern. It interferes with itself and gets a fringe pattern. Then it pulses the laser again before the surface moves much more than half an acoustic cycle, and the laser pulses again and we collect another one of those. It interferes with itself and you have two fringe patterns.

This is all done in computers. You interfere with the fringe patterns and you can make the same images of mines that I showed you, in principle. We know it does what the LDV does in terms of making images.

What we are now going to do, I think, is take it to the field, put it on a forklift, and we are going to image mines. What are we going to get? The camera has a million by a million pixels. If we can make shape work, we are going to be able to get lots of spatial resolution and be able to get shape about things is what we are hoping. Those are two things we are going to do. There are problems with it.

[Slide]

This is what the Army thinks, and they have nothing to do but to think about what might happen. It is pizza technology. If you want a pizza, what do you do? Call Dominoes, right? Tell them what you want on it. That is the way they do in the Army. They call some contractor and say, "Hey, here is what we want. We want it to go 5 km/hr, we want it to find 92% of the mines." They just order it. Nobody ever thinks about whether it will work.

This is what they think about our program, they think very positively, so positively that we are going to spend about $5 million a year over the next five to seven years doing some of this stuff.

[Slide]

Remember, I glued those accelerometers on a mine? This is a big anti-tank mine. This is an accelerometer that cost a few dollars.

This is one they give away. I glued this one on the top -- it was kind of unnerving, actually.

I glued this one on the top next to it. I glued this one on the bottom. I set this one on the ground away from the mine.

This is acceleration. What do I say? These two are close. The one on the back is clearly a lot smaller. It does not make sense to me. It seems to me the back of the mine ought to be going with the same acceleration as the front of the mine, unless the top is moving relative to the back. A lot of people believe that; in fact, if you were at the last ASA meeting, the guys from Georgia Tech are saying they are measuring resonances.

It just happens to be at the same frequency you get when you put a geophone on the ground, you do not use the laser, so I think there is something else and I do not know what it is. This was one experiment and it is not enough for me to really answer anything.

That is it.

MEDICAL APPLICATIONS OF ACOUSTICS

Lawrence A. Crum
University of Washington

DR. CRUM: Thanks very much for inviting me and giving me a chance to participate, again. I participated in the first PASS about 10 years ago. As Logan Hargrove approaches the waning years of his professional career, I wanted to acknowledge Logan and all the things that he has done for physical acoustics. I have been in a lot of different organizations, and I like the Physical Acoustic Society -- it is my favorite society -- and my favorite part of that is the Physical Acoustics Technical Committee. I think it is the most vibrant, alive, exciting group of acousticians and scientists in the world. That is not all due to Logan, but some of this has been due to Logan. He gave me my first grant and I appreciate that very much and maybe the Society is not better because I am in acoustics, but at least I am in acoustics because of some things Logan has done and I want to appreciate that and I want us all to appreciate the good things he has done for physical acoustics in the United States. [Applause]

I want to tell you something I read in the paper this morning: "Everyone knows what the process of science is about. Some reclusive egghead slaves away at some obscure subject, learning more and more about less and less until reaching the point of being so completely expert at some sliver of something or other that it is only of interest to three competitors in Berkeley, Cambridge, and Budapest." (laughter)

My wife is always telling me we are nerds and we constantly lose contact with the real world, and I say, "Oh, that's stupid, you don't understand scientists." She says, "Okay, what about the Roy Arnold story?" I say, "Okay, you're right on that one," so I have to tell you about the Roy Arnold story.

When I first went to Mississippi 20 years ago or so, Hank Bass, Roy Arnold, and Lee Bowen were the first tornado chasers. They were out there long before you saw the movie "Twister" and all that stuff. What they were trying to do was to plant a package in front of a tornado, believe it or not, an acoustic package, so that they could measure the sound and all the output of a tornado. (Hank even has a patent for a tornado detector -- when the house starts shaking, it is there.) (laughter)

At any rate, Roy Arnold was the guy who did this more than anything, and on a very low budget. He got an old pickup truck and loaded his stuff in the back and chased tornados all over Tornado Alley in the Midwest. One spring, tornado season, as soon as school was out, Roy Arnold went . off to Texas, Oklahoma, and Nebraska. He had this young undergraduate student, a young lady, who went along with him.

They were out and they got a call that there was a tornado in a particular area, so they were going to go out and deploy this package. The package had a bunch, in those days, of long microphones that they put in the ground to use as a geophone. The way they put those things in the ground to protect them from the rain and everything is they put a condom over the end of them.

They got out there and were about ready to deploy this thing and did not have any condoms. Roy jumps in the truck and the little girl gets in with him and they drive to some little country store. She is running somewhere and Roy is running somewhere when he calls out, "Jenny, I found the condoms. How many do we need?"

Jenny said, "It's going to be a long night, Roy, you'd better get a dozen of those." Roy said, "I've been tearing a bunch of those, maybe I ought to get two dozen." (laughter)

They go up the counter and, you know, this is the Bible Belt of Oklahoma, and Roy is standing there and he looks at the clerk and is about to ask, "What?", because the clerk is looking at him as if he were crazy. Jenny comes up and Roy said to her, "Gosh, I just thought of something. How are we going to keep these things from leaking?" Jenny said, "Not to worry, I've got two big rolls of duct tape." (laughter)

Anyway, you can lose contact with the real world, so I am going to talk about medical applications of acoustics.

[Slide 1]

I started this lithotripsy stuff back in Mississippi a long time ago with Charlie Church, Ron Roy, and so forth. Then we moved to Seattle and we have now what is called an NIH program project. A program project means you have grants from several different institutions. The names of the institutions are here and the names of the people involved (I will not go through those); I just wanted to tell you there are a lot of people involved.

[Slide 2]

What is lithotripsy and why do you do this? By the way, I am going to give three lectures: one on lithotripsy; one I am going to call general aspects of ultrasound; then I am going to specialize and do one on acoustic hemostasis.

In terms of lithotripsy, you get these stones. Ten percent of the population has kidney stones at some point in their lives, so there is a good chance that at least one of you in this room has a kidney stone. Ten percent of those, or 1% of the population, become symptomatic at some time, so of the almost 100 of us here, it is not unlikely that one of us in here at some time in our lives is going to have some sort of treatment for these kidney stones.

They form because you do not drink enough beer and you drink too much milk. They are basically calcium oxalate stones that form in the kidney and they usually just hang around and bang around in there and do not usually cause that much problem. If they fall off these sort of stalagtite-type formations of the kidney and pass through the system before they get big, you are okay.

But if they get so big that they cannot get out and they start blocking the ureter here, then you get into trouble, particularly if they are big enough to get in the ureter and then cannot move. Then you are really in trouble, because then you get this enormous back pressure and your kidney starts swelling up, so you have to do something about it.

When do you an x-ray you can see these things quite easily and the idea is to break them up when they are in here so that you sort of pee sand rather than getting a blockage from that kidney stone.

The techniques until about the early parts of the 1980s were to go in there and cut into the kidney and pluck them out or to run a catheter all the way up through here and go in there and pluck them out that way. You can see here the little baskets and things that you use to actually take them out.

There was a serious paper from some Japanese who actually had a device in which they ran a little piece of explosive up there and put it against the stone and lit off the explosive, literally: I heard this person talk about how many patients he had tried that on. "Oh, we don't blow up the kidney, we don't put much explosive up there," he said.

This was such a successful technique that about 80% of people who are not grossly overweight have this done now. It is the concept of extracorporeal shock wave lithotripsy, in which you set off an arc discharge of some sort, generate one humongous shock wave, and that travels through some water tank or through some coupling medium and it focuses it down on the kidney stone. It is just like taking a hammer to the stone, you break it up, and that solves the problem quite readily.

[Slide 3]

This is a real case for nonlinear acoustics, because the waveforms here are pretty humongous. Almost everybody is treated now with the Dornier Human Machine 3. People love this machine. We work with a lot of urologists; they do not want anything else. There are lots of other machines that have been made that have come and gone, but this one lasts longer than anything.

These are idealized waveforms, but we have measured these things. Basically, in a few nanoseconds (maybe 10ths of nanoseconds), but quicker than you can measure it -- there are no hydrophones now with which we can measure the rise time of that waveform there -- it goes from 0 atm to on the order of 750 to 1000 atm within 10ths of nanoseconds.

Then it decays down and goes negative here, and it is hard to determine, sometimes, how negative it gets, maybe on the order of 100 to 150 atm of negative pressure. Then it stays negative and this tail here is not being able to be resolved by most hydrophones, because you get cavitation from the hydrophone itself, so some optical-fiber hydrophones are indicating that this thing hangs negative down here for maybe 5 to 7 μsec, and then it wiggles a little bit and goes back to normal.

What happens is, when this waveform hits a stone, it breaks up into little pieces. You can usually break it up into little pieces like this with maybe 50 shocks, but because you do not want a big stone collecting in the ureter, you continue not for 50 but for 2000 to 4000 more shocks. That is a problem, because you can break up a stone into big pieces right away but you want to get them down into really small pieces, if you are a physician, to prevent liability. That is the driving algorithm for physicians, of course, is liability.

[Slide 4]

We have done a lot of experiments at the University of Indiana in Indianapolis. Peter Chang here is using one of these small handheld ultrasound devices. We use pigs that weigh about 100 to 150 pounds. We put them in this tank and we treat the kidney -- without stones, we do not put stones in there. It turns out that the damage that one gets to a kidney -- the kidney is about the same size as the human kidney -- the damage that you get you can approximate with this porcine model.

[Slide 5]

There are about 90,000 people in the United States treated with lithotriptors every year, so there have probably been millions, if not tens of millions, have been treated with these lithotriptors, and not a single person has died where they have done an immediate autopsy. A couple of people died at the very beginning because they started heart arrhythmias, but no one has gone in and removed the kidney, so no one knows, who I am aware of or the urologists are aware of, about what happens to a human kidney when treated with lithotripsy, but we know what happens to a pig kidney when it is treated with lithotripsy, and this is what happens to a pig kidney when it is treated with a lithotriptor.

I am going to show you, in this little movie sequence here, slices through that kidney, what happens. [Movie]

Now I am slicing through the kidney. These are really thin slices that are then photographed, and what you are seeing here now is the damage to the kidney. The dark areas are hematomas; that is, broken blood vessels that are releasing blood into the system. You can see massive hematomas in the area of the kidney itself.

We are slicing all the way through and you can see there is a big hematoma on the outside, a subscapsular hematoma there, and now that is starting to show here as we cut down through it. This is all blood and damaged tissue as we slice all the way down, so this is pretty massive damage to the kidney.

DR. BASS: Is the lithotripsy normally accompanied by significant urinary blood loss?

DR. CRUM: There is significant hematuria but doctors say, "Come on, I broke up that stone," and all that sort of stuff.

These people at Physical Acoustics Society meetings come to the lithotripsy sessions to hear all these anatomists who are in this program talk about the damage to the kidney in terms of the biological damage, but I am not going to talk about that. I am going to talk about the physics of it, if I can, but I wanted to impress upon you that there are significant amounts of damage to the kidney. Fifteen percent of the kidney mass is destroyed, it has become scarred, and so forth, but some people are born with only one kidney and you never know until you do an ultrasound scan or they die, so you can live easily with one kidney, it is not a problem at all, you have a lot of reserves there, so it is not a big problem. It is just that you do not want to damage a vital organ unless you have to, so we are trying to understand the mechanisms that cause biological damage and do something about them.

Even back in Mississippi, when we were starting to do this stuff, we put some pieces of chalk (pieces of plaster of Paris, really) and applied the lithotriptor for 10 shocks, 20 shocks, and so on, and you see this debridement that occurs, it just slowly eats away. This is a soft stone rather than a hard stone. A hard stone, like a marble, just shatters, but if you have a soft stone, you slowly debride or ablate it away.

Those little pits remind me, and those of you who know anything about cavitation, of cavitation, so we proposed in some applications to the NIH that this was cavitation and cavitation was the mechanism that broke up these stones.

[Slide 6]

I learned something about lithotripsy when I was on sabbatical in London in 1985, when they had the first lithotriptor in Europe, or at least in London, at Saint Thomas Hospital. I saw that and I said that has got to be cavitation. Of course, everybody who knows me knows that I see cavitation everywhere all the time.

What can cavitation do? Cavitation is the collapse of a bubble. When a bubble is expanded by an ultrasonic field and then collapsed, an enormous amount of energy can be associated with the collapse of the bubble, particularly a liquid jet that shoots at maybe the velocity of sound in the liquid against a surface, and that causes enormous damage.

Here are some original propellers, screw propellers, screws from old steamships, and look how they have been debrided by cavitation. The edges are all worn away by cavitation. It took a long time before the military learned how to solve the cavitation problem for screw propellers.

[Movie]

Now I want to show you just a high-speed movie of cavitation in a lithotriptor, and it is really going to be impressive. It is very fast. The shock wave lasts for 4 or 5 µsec, so the cavitation bubbles are formed and go away in maybe 10ths of microseconds, or perhaps even 100ths of microseconds, so you need very, very high-speed photography to see that stuff.

Right now we are trying very hard to get the resources necessary to purchase high-speed photography.

[Slide 7]

Here is a stone being held by a little device and I am going to show you, again, that I think cavitation is important in this.

By the way, when I was in London I went to visit Dornier, who was making these things, and I said, "I have the answer to all of your problems. I know what's causing stone breakup and let me tell you how to solve the problem." There were almost as many people from Dornier in my lecture. I said it was cavitation that was breaking up the stone. They all listened very politely, thanked me and they left and, as I left, I asked the guy who had hosted me, "How many people believed what I said?" He said, "Ah, probably two or three."

I think we are still right; we have not proved it to the community. Most of us in the group are going to Germany next year, after 10 years, to still try to convince them that cavitation is important.

This is a little bit closer. There is a stone. All these cavitation bubbles kind of collect on the stone and this is still running. Now watch: Some of the cavitation bubbles collapse right there -- you see that collapse. Watch in time as those bubbles collapse. See the little debris coming off? The debris comes off synchronously with the collapse of the cavitation bubbles themselves.

What happens when you have a cavitation bubble collapse is that if you have a piece of foil -- this is sort of the blast zone, which is about the size of a small cigar, inside the body. The HM-3 lithotriptor works very well because doctors get bored. They aim this thing at a stone and then you have to shoot this thing, let's say, 4000 shots, and you cannot shoot it any faster than one per second, so you can imagine, after the first 100 or 200 or maybe even 1000, you are getting kind of bored, so you go outside to have a cup of coffee.

Meanwhile, the person who is lying there getting this lithotripsy decides he wants to twist a little bit, so the stone goes out of the shock-wave focus and if it is not right at the focus, then you are missing the stone and you are not going to break up the stone.

The HM-3 has this huge blast path that is as big as a cigar, so that even a doctor can hit a kidney stone. What happens in these pits here -- this is foil -- is you see these enormous holes produced by this collapsing cavitation bubble. This shows you can tear a hole in it.

I even put a piece of brass plate about a quarter-inch thick in there and you can see the holes that were formed after just one or two shocks. Here is a blowup of that, maybe 15 microns in diameter, but enormous damage from one cavitation bubble collapse.

DR. COSTLEY: What is next to the kidney? If you move this thing, where does it hit?

DR. CRUM: Usually this stone is in the collecting volume of the kidney, which is surrounded by urine. That is very important, because urine you can make cavitation in. It is very difficult to make cavitation in blood.

[Movie]

Dahlia Sokolov is working on this particular project and she took some of these movies and I want to make sure that she gets credit. She is trying to understand how to control this cavitation. In this particular sequence she has two lithotriptor pulses shooting right at each other. We are trying to actually concentrate the cavitation only at the stone, so this is a very difficult timing thing.

This, I think, is a very interesting movie. There are two shock waves coming in. This is on a piece of foil. Watch these bubbles collapse right on the foil -- right there, ping, it is collapsing right down on the foil.

Now the bubble goes away and look at the damage sites there on the foil. When those bubbles collapse, they throw these jets right into the foil and they make holes, in a sense. If this were a piece of brass plate, you would see a big depression in the brass plate. There is an enormous amount of energy associated with the collapse of that event.

[Slide 9]

There is a lot of mathematics associated with this and I am not going to go into that, it is just mathematics. A lot of people have been trying to understand what happens for a long time. We often talk about a Rayleigh cavity. Cavitation was first looked at seriously in an analytical way by Lord Rayleigh. We talk about a Rayleigh cavity and we talk about the Rayleigh-Plesset equation after a professor at Caltech who expanded Rayleigh's equations, and so forth.

This one here is called the Gilmore-Akulichev expression, an equation that describes the radius time curve of the bubble itself. If you have a bubble and you apply some sort of acoustic field to it, the bubble is going to oscillate. If it is one short little lithotriptor-type pulse, then you can examine it, basically, by the Gilmore equation, and this is an expression for it.

Here is the acoustic field that gets inserted into the Gilmore equation. One of the other things that we have learned is that when the bubble gets big, it gets big (I will show you this radius-time curve) for a long time -- relatively -- 500 μsec, so gas diffuses into that essentially vacuum. So now there is a diffusion equation that dumps gas into the bubble itself, and we can talk about that. That happens to be an important quantity.

Again, just pictures rather than numbers, so if the waveform looks like this, there are three sort of pressure pulses radiating through the fluid.

[Slide 10]

This little picture here says that we have an ellipsoidal reflector and an arc discharge. An arc discharge goes off and you are expanding now, you are dumping energy into an expanding bubble, so there is a direct radiated ray that goes shooting out initially very quickly. It turns out we did not know until just recently from Tom Matula's work that that was important.

Then all of that energy from that expanding burning shock wave, so to speak, as that arc discharge goes off, hits this ellipsoidal reflector and then that comes out to focus and that is what you see here. But then there is a big cavitation bubble that was formed by this expanding arc discharge. That goes out, collapses about 3 msec later, sends off a shock wave that hits the ellipsoidal reflector and comes down in later.

[Slide 11]

This is work by Andy Coleman and some people in London. They are showing only two of them here. This is the principal shock wave, this is the delayed bubble-collapse shock wave, and now what happens to a bubble if you do a radius-time curve?

When this positive edge hits here, if you have a bubble that you assume is already there, and you have to assume that you have a bubble already there, because if you do not have a bubble there you are not going to get one, so what happens after a few shocks is that you have bubble nucleation sites or little bubbles, let's say, on the order of 3 microns in size. When that positive pressure hits it, you collapse that.

Then the negative pressure here starts the bubble growing and it takes off. This is on the order of maybe 5 to 7 μsec. This time here is on the order of 300, 400, 500 μsec. Once you get that cavity growing, it grows by inertia, so this is called inertial cavitation. The bubble grows by inertia, stays big a very long time -- this is a logarithmic plot -- and then, eventually, a long, long time later after the shock wave goes past, the bubble collapses, and this area here is where we think the stone is broken up.

If this bubble collapses initially due to the positive pressure and then collapses due to the expansion, there ought to be two radiated shock waves from the bubble, so there is one here, and one there. We look for this double pulse, this dual pulse, as the signature for acoustic cavitation occurring in the kidney itself.

[Slide 12]

Ron Roy, Bob Apfel, and Christy Hollin embedded passive and active cavitation detection -- this is passive cavitation detection. This has been developed mostly by Oleg Shapoznikoff and Mike Bailey at our laboratory (Mike Bailey as well as two or three other people came from Mark Hamilton's lab). This is a very inbred group, as you can see.

Robin Cleveland, Mike Bailey, and Oleg Shapoznikoff built these what we call passive cavitation detectors, so here, now, the arc discharge goes off, you focus it down like that, and now you have a small focused transducer that is looking at only one little spot here, trying to get those bubbles that collapse in that particular spot to see if you can see some evidence of cavitation.

[Slide 13]

In fact, when you have only one passive cavitation detector, you get everything along the path that goes into the hydrophone, so what they have done now is to have two focused systems. These are high geometrical gain, so the volume they are looking at you see there. One focuses here and this one has a focus there, this one has a focus there. All three of them are confocal, so these cavitation detectors can look at only bubbles in this particular region of the kidney.

What we want to do with that thing is to scan through the kidney and see where the cavitation is occurring and see if the cavitation is occurring at those sites where there is tissue damage, so this is a device, a very crude device, in a sense, for determining a correlation between cavitation and tissue damage.

[Slide 14]

Does it work? It works pretty well. When Charlie Church first took a simple Rayleigh-Plesset equation and applied a waveform that went from zero up to 1000 bars in a few microseconds and then gone to a negative, 150 bars, in another 2.5 μsec and said, "Now I can tell what a bubble is doing," I said that is computer fantasy and do not tell me anything but that.

But Charlie stuck to his guns and now everybody sort of believes this stuff and there is enough evidence that one should believe it.

Here is what you calculate for that particular waveform. Here the bubble collapses. It grows and then it collapses again, so this is the calculated bubble radius. This is the calculated acoustic emission from this bubble now, and here is what you measure from that passive cavitation detector.

What is really interesting, you get the time right. You sort of get the amplitudes right, so this is now a relatively believable detection device.

[Slide 15]

You see this double pulse. This is cavitation growth and collapse. Now you can change the voltage on the arc discharge, 18 kV, 20 kV, 22 kV, 24 kV. This is how much arc discharge. The physicians, who are in a hurry, want to crank this thing up to 24 kV and blow that stone out of the water, so to speak.

It is not good to work here, though, because you are getting a lot of massive cavitation and that is probably associated with tissue damage, and we are trying to see if we cannot move this thing back into the lower regions and still get cavitation stone breakup.

[Slide 16]

This does not show very well, it is a complicated figure. If you remember the movie, let me try to explain this whole thing here; this is important. Here is that waveform, again, where you have the bubble collapsing initially and then it grows and collapses again and you get some emissions. Here it looks as if you even get a rebound, or something like that.

Dahlia and Mike have filmed this cavitation field, and in a free field with no stone, you see here at this particular point a clear field, and now you get some indication of bubbles associated with the initial collapse starting to grow. Here is the growth region, you see all these bubbles here, then a collapse, and when it collapses all those bubbles should go to practically nothing, so the field clears up a little bit. There is some rebound (it is kind of hard to see the rebound) here and then it clears down, again. So this is no stone.

Here is what is interesting, and I showed you that movie so that you would believe that, here is what happens if you have a stone in there. Now, instead of having to film what is going on inside the kidney, you can use this acoustic passive cavitation detector to do the inverse problem and guess, or have some measure of what is going on inside the kidney.

Now you notice, this is the time frame here for collapse (we call this the silent time). The silent time has been expanded considerably, because look what happens if you have a stone in there. When the bubbles grow, they all tend to get collected on the stone itself. There are things called Bjorknes forces that tell you that there are forces between bubbles and these bubbles probably, because there is sound reflected off the interface here, tend to get collected on the stone itself, and now this little bump there is a big bubble or a cluster of bubbles.

Now the whole cluster collapses, and cavitation cluster collapse, if we can say that, has been studied for a long time. When you do high-speed movies of cavitation on a propeller, you see sheet cavitation and cluster collapse and you can imagine, instead of having one bubble collapsing, if you have a whole bunch of bubbles collapsing, you would have a more violent collapse, and you see that is the sort of thing that we have here.

I am sure you can think yourselves of ways that you might be able to perhaps control cavitation to make this thing work a little bit better.

Charlie Church and a guy I was working with when I was on sabbatical in London measured the waveforms of various lithotriptors and because different companies have to have different intellectual property, they did them differently. Dornier had this arc discharge. Siemens has a different way of generating the shock wave. They run a huge current through a big plate that causes the plate to expand. You have this big plain wave coming out and they have an acoustic lens that focuses this down.

The Wolf Piezolith has, instead of an arc discharge, a whole bunch of transducers. EDAP had the same sort of thing. These two groups have gone defunct, I think, but, nonetheless, they could control the waveform through piezotransducers of some sort.

What intrigued me was the whole sequence, that if you made the cavitation bubble grow, then what collapsed the bubble? Atmospheric pressure. Surface tension is not important, atmospheric pressure collapsed the bubble, so in one of my sort of insights of genius I said to myself, aha, I now know how to solve this whole problem.

Instead of having the waveform first that crushes the bubble and then the bubble grows during the negative portion, why not just turn this thing around? Interestingly enough, EDAP did. Their system did not work, but they were trying to drive this thing at 60 Hz, because doctors do not want to wait for an hour or two to do this treatment; they want to get it done in 10 minutes, so they ran this thing at 60 Hz, but it did not work, and I will explain later why this thing did not work at 60 Hz.

I had the idea that if you turned this thing around and made a waveform with, instead of a positive pressure and then a negative pressure, the negative pressure first and then the positive pressure this thing would work.

[Slide 17]

I proposed that in a proposal and David Blackstock was involved as a reviewer on some of this. He had a student named Mike Bailey and Mike said, "That's a great idea, I want to do that for my dissertation project," working for David Blackstock down in Texas.

David said, "I don't know anything about lithotripsy," and Mike Bailey said, "I'll go up and work with Larry Crum." Anyway, Mike Bailey did this thing. It was a huge mistake for Mike Bailey, a wonderful success story for me but a huge mistake for Mike Bailey. You know, this makes sense, and anything that makes sense initially usually is wrong, I think.

This was my idea and how do you turn this thing around?

[Slide 18]

Mike had the idea that instead of, when the shock waves goes off and you collapse all this stuff back here by a reflection off of an interface, having a rigid reflection, you have pressure-release reflection, so he made a styrofoam reflector.

Now when the shock waves comes out and reflects, it comes out negative leading and then positive. That seems like a good idea, doesn't it? All my ideas always seem like good ideas. But if you take this thing and fire 100 shots at a stone, it breaks up like that. If you take this thing, the negative leading edge, and fire 100 shots at this stone, it does this. A wonderful success story. (laughter)

Actually, usually when something does not go right, it gives you a chance to learn something, and sometimes just learning something is useful in itself. So what went wrong here?

[Slide 19]

Probably cavitation went wrong. Here is what you have when you have a rigid reflector, shooting now, instead of putting the foil perpendicular to the beam you put it parallel to the beam, so the beam is coming in like that, and this is showing you the blast path of the focus of the lithotriptor. This is pretty big, this is 2 cm across here, so this is maybe 10 cm along there, so this is a big blast path.

These are all cavitation bubbles that have occurred. If you do the same thing with a pressure-release reflector, you still see a little bit of cavitation, but very little. What is interesting is that if you go to a pig, take the pressure-release reflector and shoot 4000 shots at the stone in a pig, you have two things happen. One is, you do not break up the stone -- that is, of course, bad -

- but you do not damage the kidney, so that is good. Do no harm. That is the first rule for a physician, do no harm. Do not necessarily fix the situation, but do no harm. (laughter)

So why doesn't this thing work? It took us some time to figure out why it did not work. If you have the negative leading edge, then the bubble starts growing very dramatically here. Then, if you have this positive pressure, it collapses, and look at this collapse. It is a fantastically enormous collapse.If you believe the theory, this thing would be such a violent collapse it would blow the kidney stone right out of the water.

The problem is that you are all assuming sphericity of the bubble behavior, so you can imagine the bubble grows up and then you suddenly hit it with this humongous positive pressure. Well, the bubble is no longer going to behave like a nice spherical cavity; it is going to collapse.

I think what happens with this positive pressure here is it just shreds the bubble and you get nothing but a bunch of little bubbles, or something like that, so you lose the comparison between theory and experiment. You no longer can model it by this waveform here. I think what is happening in here is that you destroy the bubbles with that positive pressure, so anything down here that drives the bubbles again is probably way out of phase and it just does not work.

[Slide 20]

Another thing that happens that is quite intriguing is that if you have a positive pressure like this and the bubble gets really big, it is a vacuum and it stays a vacuum for 500 μsec, so gas, if you now go in and look at the diffusion equation, will diffuse into the bubble, so when you start out with a 3-micron bubble, after one shot, one shock wave goes through you should have a 16-micron bubble. You pump a lot of gas into that cavity, so we eventually learned how to interface this diffusion equation with the Rayleigh-Plesset equation and now interesting things happen.

[Slide 21]

If you change the pulse repetition frequency -- I told you about EDAP having a system where they were going to fire at 60 Hz and it was going to solve a lot of physician time problems -- here is what happens if you fire it at a slow rate (I do not know what the slow rate is). This is a high-speed movie of the cavitation field and you see you get bubbles and they collapse and go away.

If you fire it at 3 Hz, not 60 Hz, look how many more bubbles you get. Why do you have a lot more bubbles here than you do here? I think that when these bubbles grow and collapse you pump air into those bubbles and now you have got a 60-micron bubble that is sitting there in the urine, in the kidney, and it is slowly dissolving away.

Now the next shock wave comes through before it dissolves away and you have a whole bunch more nucleation sites and a whole bunch of bubbles, so what happens is that you are reseeding the fluid with nucleation sites. You can see that after a while you are trying to shoot through a bubble cloud.

We learned in underwater acoustics that you cannot propagate through the surf zone, because every time you have a breaking wave you get huge clouds of bubbles. That cloud of bubbles looks like a wall to the acoustic field and you cannot propagate through it. I think that is the explanation for the EDAP failure in cycle rates.

We now started looking at what is the optimal cycle rate. Could you cycle it at 10 Hz, could you cycle at half-a-hertz? What is the optimal cycle rate, and let's look at the behavior of these bubbles.

This is at 0 atm of applied pressure, so here is a pressure chamber -- this is just a Coke bottle -- where we can apply 3 atm of pressure. Look at what happens if we scan this thing now with a diagnostic ultrasound system. A diagnostic ultrasound system gives you an image of the whole system. If you apply no ambient pressure (this is 1 atm of ambient pressure), then you can see the residue after the shock wave goes through, but if you apply 3 atm of ambient pressure, by the time the ultrasound scans all the way through and forms an image, there is practically nothing there, which means that if you apply overpressure, you can get rid of the daughter bubbles that formed from the previous shock wave.

[Slide 22]

Oleg Shapoznikoff and Vera Koklova now sort of interfaced this whole bubble dynamics thing together, so they have the shock wave here, they get the positive compression and the negative pressure comes up and makes the bubble grow. It collapses and now they have got diffusion in the equation, and it slowly diffuses under ambient pressure. This is 10 orders of magnitude in time here, so you can put the whole thing on one figure, so this is the lifetime of a bubble inside a kidney.

It grows and collapses, and what they found was that if you have overpressures, very modest overpressures, these residue bubbles -- remember, the bubble starts here, grows by diffusion to this point, and then collapses if you have ambient pressure on the thing. If you have atmospheric pressure, the bubble dissolution time is on the order of 60 seconds, which means it takes a long time before those bubbles dissolve.

If you put just a few atmospheres of ambient pressure, you can run that time, so you can cycle this thing much faster if you could put people in hyperbaric chambers. You could really shoot fast. We would propose that to lithotriptor manufacturers.

[Slide 23]

Here is a little bit better demonstration of that. Here is this Coke bottle, something like this, and this is a scan with no overpressure. You can see all of the sound scattering from the nuclei, from the bubbles that are formed. You apply 1 atm of overpressure -- this is indicating the scan of the diagnostic ultrasound system -- and you see it is completely clear.

We can do things with overpressure to control the daughter nuclei that allows you to recycle and obviously have some input on tissue damage, because if you get these bubbles in the blood, and now you are constantly firing these lithotriptor pulses at bubbles in the blood, that is probably where you get the damage.

[Slide 24]

Here is a higher pressure chamber. We decided if you could do good things with 3 atm of pressure, what could you do with 100 atm of pressure. Mike Bailey and a few other guys made a real pressure chamber here, where they have windows, pretty good acoustic windows, and you can now apply up to 100 atm of pressure.

DR. SABATIER: In one of your earlier slides you showed a person lying on a table with a transducer coupled to him and you had one on the top that showed him in the water.

DR. CRUM: Yes, they have different stages. In the HM-3 you just lie in the water, that is your coupling medium.

DR. SABATIER: My question is, is that enough increase in pressure when you are below the water?

DR. CRUM: I do not think so, that is 1 m, or something like that. You have got to transmit that pressure inside the kidney itself, so if you really had a lot of pressure in your bladder, maybe that would help. Drink three or four beers before you went in, that might help. It is not enough.

Here is an interesting thing. This is pulse-repetition frequency and overpressure, so even at high pulse-repetition frequencies, up to 3 Hz now, if you apply overpressure, you have no residual bubbles after one shot. This, of course, is a little too much of a fantasy to think about hyperbaric chambers but, nonetheless, we wanted to push this thing in trying to understand what was going on.

[Slide 25]

One of the things you can do is to calculate that if the bubble is growing up like this and then it collapses, there is still the concept -- you cannot forget this concept, or at least I could not -- of if you had a strong external pressure that bubble would collapse more violently, so if you can calculate, in a sense, the normalized peak radiated pressure, that is, the violence of the bubble collapse, as a function of overpressure in terms of atmospheric pressure, that is, if you apply this overpressure, doesn't the bubble collapse more violently?

It does, and if you calculate how much it collapses, what the effect of that is, you get this almost factor of eight increase in the violence of the collapse if you apply 30 or 40 atm of ambient pressure -- not that you are going to do this to humans, but there might be something that you could learn from this whole process here.

That is the calculation. Does it work? In foil you see that there is some increase as you apply external ambient pressure. The big issue here, the big quandary in this whole thing, is what is the mechanism that breaks up the stone? Most of you, because you are listening to me and you know that I am a brilliant person, a scientist, and everything, are saying, "Oh, yes, he's right, it's cavitation breaking it up." There is not a single person, probably, at Siemens or Dornier who believes that cavitation is involved.

They think that it is the shock wave itself that is breaking it up. I argued, "Look, Mike Bailey made this pressure-release reflector, he's got the same amount of positive pressure, the same amount of negative pressure, the same energy in the shock wave, and he can't break up the stone."

"Well, maybe your rise time isn't quite right," or something like that. I said, "If I can completely get rid of cavitation, would you believe?" That was said in a sort of rhetorical sense. "If I got completely rid of cavitation, would you agree that cavitation is important in breaking up the stone?" That was a good rhetorical question, so that was sort of a goal.

If you apply lots of overpressure, eventually that negative pressure cannot produce a bubble, so, therefore, you would not get any cavitation and you should not break up the stone. Well, that is what happens. If you apply elevated pressures, extremely high values of elevated pressures, you cannot break up the stone, but if you just reduce the pressure to 10 atm, or something like that, you get even enhanced breakup of the stone, all consistent with the theory.

DR. MARSTON: There is a counter to that argument, however. If the mechanism is falling, which is one of the alternative mechanisms related to the reflection of the shock wave from the back of the stone, independent of the length of the shock wave, one could argue that that it would also be inhibited by the static pressure.

DR. CRUM: Yes. I am having one of these senior moments where I forget the name of the guy who always has a counter to all of my arguments -- it is my wife, but there is also someone who is a scientist, and maybe I will remember his name and maybe I will give you the answer at the end about why I may, again, be wrong.

[Slide 26]

This is interesting, just as an aspect. If you apply all this high ambient pressure, why wouldn't you destroy all the nuclei and, therefore, you would never get cavitation? Mike Bailey, either by hook or by crook or by intention, actually had a little crack in a glass plate and he applied zero ambient pressure and you see the cavitation, but if you apply high pressure and scan it, you still get some cavitation, because that crack traps gas.

Even if you applied high pressure to the stone, you would still have little nuclei trapped inside the stone and we think that this is an important component of the whole process.

[Slide 27]

To summarize this, Mike and Dahlia have been trying to find ways to actually take physical acoustics and do something with it that Siemens could use, or Dornier could use, to make a better lithotriptor. That is an important goal for us all, to have some impact of our science.

One of Dahlia's projects was to see if she could goose the bubble at exactly the right time, so here is the bubble growth and collapse. Let's suppose you hit the bubble at exactly the right sequence in its collapse. If you hit it up here, you are probably going to destroy the bubble, because it is trying to grow and nothing happens.This is foil and there is almost no damage at all to the foil when you hit the bubble with a second pulse at this particular stage.

If you hit the bubble at this particular stage with a second pulse, you get lots of cavitation and it is spread around all over the place.

If you hit it right here, it is too late, so all the cavitation is generated by these individual interactions of these two shock waves. [Movie]

Let me show you, again, the sort of thing that Dahlia is trying to do and trying to understand. We think this might be an important aspect of this. What Dahlia is doing is firing a shock wave in like this, and one in like this, and trying to time them so they overlap right at the stone itself.

Now, back here, if this were just one shock wave coming through, you would have cavitation bubbles about like that, but because as the shock wave goes through, it destroys all these bubbles here and, as this one goes through, it destroys all the bubbles there, you get behavior only here in the middle, so you have concentrated the cavitation at one particular spot.

What is interesting is that you get what she calls a butterfly pattern here of the interaction of these two shock waves coming in here in the field itself, and she has been spending a lot of time trying to understand that. This has a lot of promise, we think, for concentrating the cavitation at one particular site.

You can see now the area of concentration of the cavitation is only in a very small spot and we think that this is something that would be very useful, although we cannot envision someone being in a tank with lithotriptor pulses coming in from two different angles, but you can envision ways of doing it.

[Slide 28]

I was going to show you, again, the foil stuff, but I want to end now, because time is running outt, with a lead-in to Tom's stuff -- maybe he will not talk about this -- but at least it is a lead-in to some of the things he is doing. We wanted to find out whether you get sonoluminescence from a person during lithotripsy -- that seems kind of interesting.

This is a true story. We had a site review from the NIH; in a program project they bring all these people in and we have to give a presentation and then they give us a score to determine whether they are going to fund us for another five years. It is a big project, so this site visit is a very important thing.

Brad Sturdevant, whom maybe some of you know, from Caltech, was there to give the presentation and the very day that he was to give a presentation he had a kidney stone drop down in the ureter and block it up, during the site-visit time, so he is in the hospital and cannot give the presentation.

Jim Lingaman, who is our lead urologist and has treated more kidney-stone patients than anybody, probably, in the world (they were treating 6000 patients a year at the Indianapolis

Center for Stone Disease), runs a catheter all the way up into the ureter and he films all this stuff. He is up in there and he is reaching in, trying to grab the stone, and we are watching some of this and we say, "God, this is fantastic, you're looking at this through an optical fiber, right?" He says, "Yes, yes." "Let's hit it with a lithotriptor, put a PMT on the back of it, and see if we are getting sonoluminescence out of this guy." (laughter)

He said, "No, you haven't got the protocol, you can't do that." We actually had everything all set. Robin Cleveland was running out, trying to get the PMT, and Mike was doing the optic; we really had a fun time.

But you really can get sonoluminescence out of this whole system, so here are the shock waves coming out, the radiated shock waves coming out, and here is the sonoluminescence and they correlate very readily and, interestingly enough, you sometimes have more light coming out of the compression than you do out of the growth and final collapse.

I am finished with that sort of thing and now maybe a little bit of time for questions.

DR. GARRETT: What makes you think it is sonoluminescence and not triboluminescence?

DR. CRUM: Well, if you did it in just ordinary water with the shock wave, you get sonoluminescence, so if you put a stone in there, you tend to get more sonoluminescence, so maybe there is some triboluminescence, but if you do it without a stone, you still see the sonoluminescence. In fact, I am not so sure what the answers are to that, but I think you would probably get even more, because you tend to get more spherical collapses when you do not have a boundary in there, so you do get sonoluminescence out of it.

One of the things that Tom is trying to do is to levitate a single bubble and hit it with two lithotriptors so that we can do single-bubble sonoluminescence at, instead of 1 atm, 100 atm, so we really want to goose the thing.

MR. GLADDEN: That last video that you showed with the butterfly pattern, what was the length scale on that?

MS. SOKOLOV: The vertical dimension on that was 2 cm.

DR. CRUM: Eisenminger, I wanted to talk about Eisenminger. Let me answer your question and then you can ask another. Eisenminger is this brilliant guy, I think he is from Stuttgart, and I give a paper and then I say, "Wolfgang, what do you think of that?" He says, "Yes, but

-- you haven't thought about this," and always I had not thought about that, but Eisenminger, as some of you know, is this very brilliant guy, so I showed these data at Berlin and said, "See, it has to be cavitation, because we apply atmospheres of pressure, 50-60 atm of pressure, and we cannot break the stone.

"If we flip the waveform, just change it, we get the same amount of energy, we can't break the stone. Now, tell me how spallation, and so forth, could cause the stone to break up and how come cavitation is not important?"

Well, here is his answer. He says that he thinks that the stone breaks up by microfractures and spallation, so the shock wave comes in, hits the back, reflects off, and now you develop these nice little microfractures and, after a while, of course, it is like taking a stone -- if you have ever seen a stonecutter cut a brick, what he will do is he will tap, tap, tap, and then hit it really hard and it breaks along that area where he made the microfractures.

What he says is happening is that the shock wave is causing microfractures and then eventually it just breaks apart under spallation. The reason -- this is probably a good reason -- that when you apply high overpressure you do not get any stone breakup is that you compress everything down and now the microfractures do not break up, because everything is under compression. It is like prestressed concrete.

Maybe these guys are right, after all, and maybe the shock wave itself is extremely important in breaking up the stone, but obviously cavitation has some role, too.

DR. MARSTON: On your study of the effect of overpressure on the.... , could you say something about what the gas content was in the liquid in your overpressure model?

DR. CRUM: In my next presentation I am going to talk about contrast agents, and we used perfluorocarbon, these gases that have a very low solubility, but in urine you have probably got mostly air, so the diffusion is mostly air rather than CO_2 or anything else like that, so we do the diffusion equation with air in there. Is that your question?

DR. MARSTON: But it is essentially when you overpressurize the urine, is the urine saturated at the overpressure value?

DR. CRUM: It depends on what the time is, what time you apply there, and all that sort of stuff, so we do not want to do that. We do not want to supersaturate the fluid. If you apply ambient pressure with a gas ullage above it, then you will eventually supersaturate it and the equilibrium concentration of gas in the liquid will be 3 atm rather than 0 atm, since we try to do it fast enough that the equilibrium concentration is at 2 atm and the ambient pressure at 3 atm.The bubble in the compressed stage is actually saturated with respect to the gas, so it would normally diffuse out. We want to dissolve those bubbles. The scheme here is that what we are trying to do, and I ended too quickly, I did not have time for a summary here, is to reduce the damage. We are pretty sure we do not have any problem in breaking up the stone by cavitation, but if we have all those little bubbles around -- that first slide I showed you was all this damage to the kidney -- we think these are bubbles that are going into the capillaries and the cavitation is going by and that is causing damage to the capillaries and you are getting hematomas.

We think that if we could somehow or other do two things, prevent those cavitation bubbles from growing inside the capillaries we would not get damage, and the other thing is we are thinking of putting in some sort of liquid, vitamin C, or something like that, that would absorb the free radicals that can cause the potential bioeffects of the cavitation byproducts, these free radicals and everything, from causing damage, and maybe even do something with the mechanical effects.

DR. SPARROW: You are using this for kidney stones. I heard a while back that people were trying to do this with gallstones. Was that successful?

DR. CRUM: Yes, the reason people made these things was for gallstones. The reason they made them for gallstones is that the prevalence of bile disease is 10 times that of kidney disease, but it turns out now, with these endoscopic microscopic techniques, you just stick a little needle in there and you go in and you suck out the gallstone, because it is not a hard stone and you can essentially put in some fluid and dissolve it and suck it out. Almost all bile stones are treated with laparascopic techniques now and lithotriptors are not used for gallstones -- unfortunately.

MR. PORTER: Has anyone else run a lithotriptor at a higher PRF?

DR. CRUM: This guy, Delius, had one that was running at 60 Hz and he found out that he could not only not break up the stone but had orders of magnitude higher damage to the kidney when he went to these high pulse-repetition rates, because, I think, you have got all these bubbles in there and you have all these cavitation effects of the bubbles. He never could get to the stone.

That is why they cannot cycle it faster. The maximum rate they will approve it now is, I think, 2 Hz, the maximum rate they will permit.

DR. WAXLER: Have you tried any of these techniques to reduce bubble formation or cavitation on animal kidneys?

DR. CRUM: Yes, it would be nice to pull all the urine out and de-gas it, put it back in, and so forth, but you can take de-gassed water and as soon as you hit 10 to 20 lithotriptor shocks you have taken all that gas that was in solution and you put it into those bubbles and now you have a faster dissolution rate but you cannot take this fluid out very easily.

What is interesting is your body does not have many nucleation sites in it; otherwise you would never be able to go scuba diving. In 40 or 50% of the cases in which you give a lithotriptor treatment the first thing they do, often, is run a catheter in there and put in x-ray contrast agents so that they can see if you have any stones in your ureter. If you have just kidney stones and you are hurting, they do not want any stones in your ureter, so they often do that.

What do you think they squirt in there? It is phosphate-buffered saline, not de-gassed phosphate-buffered saline. They just put very gassy water into the collecting volume of your kidney. It is hard to control the fluid inside someone's kidney.

Let's take a break.

DR. CRUM: I want to talk to you now about mostly medical ultrasound and I wanted to also tell you that I am just a spokesperson for a lot of this work that is going on. We are now trying to build a program that is called the Center for Industrial and Medical Ultrasound at the University of Washington. We are mostly trying to get involved in technology transfer.

The applied physics lab is about a 300-person laboratory that has been around for 55 years and, except for the director, not a single person has tenure, so we have to raise money; we are all on soft money. As a consequence, our distinction from an ordinary department is that we try to do technology transfer in the Seattle region, which is a very high-tech region with lots of start-up companies -- it was number one or number two in the United States in terms of start-up companies last year.

I am going to emphasize a bit of the technology-transfer aspect of this and tell you of our attempts to actually take some of this 6.1-type research, basic research, and actually spin off some companies. That is the motivation for some of the work that we are actually doing here.

We are involved in diagnostic ultrasound and therapeutic ultrasound, and here are some of the devices that we already have. Here are the people who are involved in the Center and the applied physics labs, where a lot of us are, but there are 11 other departments. We have 20 different companies that we are involved with right now in some way, either as SBIRs or dual-use projects, called dual-use science and technology. There are a lot of people involved and I am particularly a spokesperson for the kinds of things we are doing.

We are interested in trauma, and this is trauma. This is what happens if you are not careful when you are riding your motor bike down the road, you have all kinds of trauma that develops and we are trying to treat trauma.

[Slide 29]

One of the things that we are trying to do is to solve a problem in what we call combat casualty care. Most of the money that we are getting for our programs come from the combat casualty care divisions of the Navy and the Army. These are at the 6.3 level in most cases. We got one project from DARPA and we have one project now from the Navy, another one that is in the process of being funded from the Army and another proposal in to the Navy. They are all at these 6.2 and 6.3 levels.

We are working with a lot of companies and I want to walk through some of the projects that we are doing. In the next hour I am going to focus on acoustic hemostasis. The largest market share of diagnostic ultrasound equipment in the world is Advanced Technology Labs, ATL, which was purchased by Phillips not too long ago. This is the kind of stuff that you can with ultrasound.

I would call that a picture of a fetus, Pat Buchanan would call that a baby, but, nonetheless, you can see that you can have remarkable acuity now in determining the imaging of ultrasound with these systems.

These are, by the way, gallstones. With a device like this you certainly can determine whether somebody has gallstones. There has been an incredible advance in this technology. It is about a $3 billion-a-year business. ATL's headquarters is in Seattle, and Siemens' headquarters are in Seattle. Together, they have more than 50% of the market share of diagnostic ultrasound.

[Slide 30]

There were lots of people who were involved in trying to build devices that would solve some of the problems in combat casualty care. Imagine that you were out in Somalia or Kosovo and somebody got shot. If you tear a major blood vessel, then how do you determine whether this person is bleeding to death or he just has a gallstone? How do you do triage on the battlefield?

The way you do triage on the battlefield is that you take that HDI-5000, which weighs about 300 pounds, costs about $300,000, is about as big as a refrigerator, and you apply advances in ASIC technology and you put it all into a small thing. I wrote a proposal to DARPA under the Dual-Use Science and Technology Program to work with ATL and we got $14.7 million to take all of this stuff and put it into a single unit and make a handheld diagnostic ultrasound unit.

ATL then spun off a company called SonoSite, and SonoSite is making these things. SonoSite's income is approaching $100 million a year from just making these things.This was a success story for DARPA.

We did not get a single dollar, a single intellectual property equity share, out of this whole thing, which was, I think, a real mistake on our part in terms of learning how to do that, so one of the things that I think you young students now, maybe even you young scientists, have to learn to do, is to do technology and get a share of this technology for the government or for your university or for yourselves. That is going to be, I think, the future of science and technology, is learning how to build these new companies and make them successful.

One of the things that we were trying to do is to take a device like this, you can make it wireless, so you can radiate over here this image. It goes up to a satellite and comes right down to a radiologist -- this is our resident radiologist, Steve Carter -- and he can communicate wirelessly now and direct the scan. Steve can communicate with me wirelessly and I can do scanning like that, so maybe on the battlefield we could do that.

That is one of the things we are trying to do in terms of what we call telemedicine and remote emergency medicine.

[Slide 31]

I want to take four or five slides to show you some other things that some of the graduate students are doing at the basic science level. You can see from the basic science how we have a technological transfer potential, we think.

Tyrone Porter is doing things like this. What he wants to do is to understand how ultrasound can actually take drugs and transport them into the cells. It is called sonoporation. It is a big topic right now, because with a little bit of ultrasound, not very high intensity, and a little bit of drug, you can take that drug and stick it right inside a cell. This could be a toxic drug that you could apply to a specific site, so this is called site-specific ultrasound-activated drug delivery. You could imagine you could do this to the prostate or to the pancreas.

Chemotherapy is a vicious thing to do to people. If you could do chemotherapy in only certain areas, it would be a very important thing. Tyrone and Connie Kwok, working together, are trying to understand how you get that ultrasound to increase the transport across the boundary like that.

This is some of Connie's work. What Connie found was that if you take a certain polymer, and they make these "smart polymers," they call them, at the University of Washington in the bioengineering department, and you take a drug (in this case it is an antibiotic) and put it in a sponge, or something like that, a reservoir, and then you put a coating of this polymer on top of it and apply ultrasound, you can cause the drug to come out, so you have sonoactivated poration, that is, you release the drug.

With this antibiotic here, if you have this red line here uncoated, if you put it inside the sponge and you apply ultrasound, then when you have the ultrasound on, you get this release of the drug. When you turn the ultrasound off, it comes down, and now it is still leaking out. This is not good.

But if you apply this polymer and you treat it (not with ultrasound) with some sort of incubation scheme to harden up the carbon chains, to lengthen the carbon chains, now look what happens -- it is the green line. You turn the ultrasound on, you get release of the drug, and it almost goes back to ambient or no release. Whenever you turn the ultrasound on, you release the drug; turn the ultrasound off, you do not release the drug.

If you can get this release rate high enough and you can get a feedback mechanism, you can do this with insulin, so now there is insulin release with these things. What we have in mind is that you would embed this insulin sponge, apply ultrasound, release the insulin, and you have some feedback mechanism looking a the glucose level inside the body.

[Slide 32]

Another thing that is very important in terms of applications of ultrasound is the blood-brain barrier. If you get a tumor in the brain, you are almost a goner for a variety of reasons, one of which is you cannot do chemotherapy on the brain because of the blood-brain barrier.

Here is a capillary. Here is brain tissue. Your brain is protected so that drugs, viruses, and bacteria do not get into your brain, they do not get across that blood-brain barrier. Pierre Murad and Dahlia worked on this stuff for awhile.

Here is a rat. You take off the skull, that is the brain. You apply ultrasound -- it is laser-guided, which just means you know where to apply it -- to the brain and at the same time injected into the bloodstream is a dye. This is basic research. When you apply the ultrasound you can see that you have the penetration of the dye into the gray matter of the brain itself.

So you open the blood-barrier. You can do that in lots of different ways, so could you do it with ultrasound and what is the mechanism? We have spent a lot of time trying to find the mechanism. It turns out that -- right there -- is what is sometimes called a tight junction. If you look right in there -- I cannot tell this and you cannot tell this -- but a good neurophysiologist says, "Aha, I see that the tight junction is open," and you could actually have passages of things in like that.

It turns out that when you apply ultrasound here, this becomes available for the transport of molecules as large as 70,000 Da for up to 96 hours. This would be an ideal thing. You could do chemotherapy on the brain if you could apply the ultrasound in at the right intensities, and so forth.

[Slide 33]

Here is another company that we are involved in starting, and we do a lot of SBIRs with that. A lot of people, particularly older people who do not exercise, get clots in the veins in their legs and those become very painful and can become dangerous, so how do you remove those clots?

One of the ways of doing it is to put ultrasound on the tip of a catheter. Here is a catheter, it has a little transducer on it here. You run the catheter down into the vein in the leg and you squirt out a thrombolytic drug. A thrombus is a clot and lytic means to dissolve, so you put in a clot dissolvant, aspirin, or something like that. These are urokinase, streptokinase, and tPA. You take this catheter, you run it right into the clot, squirt out the drug, turn on the ultrasound.

Here is an example. Here are two veins that are blocked. You put a catheter in both sides, you turn the ultrasound on in this one and it releases in just a few minutes. You do not turn the ultrasound on this one and you have the drug coming out and it does not release.

These are examples of the fact that ultrasound-enhanced drug delivery works very well. It turns out this company was going along fine, with about 70 employees, had spent about $25 million, and it turns out that you can dissolve a clot in the leg in, instead of four hours, two hours, and you would think that would be a good sales things, but physicians do not want to wait even two hours. They say, "What the hell, just put a drip into the arm, I'll go home and come back in the morning, so it will dissolve in four hours. Why should I be involved to save two hours?" The whole company almost went broke until now, when it is going back to an even more challenging problem: What if you have a clot in the brain? You do not have two hours.

Now they have made this thing small enough to get into clots in the brain when you have a stroke and now they have had great success. They have recruited only five patients but in each case they have been able to dissolve those clots very rapidly.

[Slide 34]

One of the things that you can do in terms of ultrasound, you see all these beautiful images, but the biggest challenge in diagnostic medicine right now is to determine how extensive your heart attack is. If you have a heart attack, how bad is the damage to the myocardial tissue?

You can do ultrasound and you can see the blood vessels going into the heart. Could you do myocardial perfusion; that is, how much blood is going to that damaged area of the heart? That is the sort of Holy Grail or goal of ultrasound nowadays. The idea, then, is that if you can use some technique for enhancing flow -- this is contrast agents, small stabilized microbubbles, and some very nice new technology called harmonic imaging.

If you put a bunch of bubbles into the kidney -- this is the kidney now -- and you look at the enhanced scatter from these bubbles, you can actually look at all the blood flow through the kidney. Now, if you could look at the blood flow through myocardial tissue of the heart, it would be even better.

I show this picture to show we went in and damaged, we punctured holes in the kidney there, preventing perfusion to that particular area, and we pointed out to our military sponsors that we could determine damage to kidneys under combat casualty rather than the more commercial myocardial perfusion.

[Slide 35]

In a lot of cases we are interested in therapy. It means we want to have a biological effect. In some cases you do not want to have a biological effect. If I am scanning the heart to see if there is myocardial perfusion, I do not want to violate the first law of Hippocrates: Do no harm.

Here is what Andy Braman, a new person, has looked at. He has looked at the damage that contrast agents applied with diagnostic ultrasound intensities would cause. He has scored these sorts of things here. At this particular intensity you have complete destruction of a layer of cells in a particular area.

Here you have none.

[Slide 36]

I want to show you his results that show the effect of these contrast agents. Remember, now, almost every new diagnostic ultrasound machine that is sold for therapy has to have tissue harmonic imaging and harmonic imaging systems, because everybody wants to use contrast agents, which are stabilized bubbles. Why wouldn't you get cavitation from this and why wouldn't you cause damage?

The whole bioeffects community is now excited, again, because we can now look at the effect of cavitation due to diagnostic ultrasound machines.

What Andy was able to show in terms of this cell damage here, a damage score, is that if you do not use contrast agents -- this is Albunex -- then you can get into the pressure regimes here of 1.5 at 15 atm before you see much of any damage.

As soon as you use contrast agents, Albunix, even at these relatively low values, you can start to see damage, so you have to consider benefit versus risk. If you are trying to determine if somebody is going to die from a heart attack, you can cause a little damage to the tissue in the heart, but if you are just looking for whether a woman is pregnant or not, or if you are trying to track an egg moving down the Fallopian tube, you sure as heck do not want to be causing that kind of damage.

[Slide 37]

One of the things that Wen Chen is working on here is these contrast agents. He is working with a couple of companies, ECPS and Point Biomedical, in trying to understand not only the role of the contrast agent in terms of ultrasound-enhanced imaging but also the bioeffects.

This is kind of the life cycle of a contrast agent, according to his cartoon. Optison is a bubble that is filled full of octafluropropane. It has an albumin shell and it is a relatively thick shell. When you apply ultrasound, even at diagnostic ultrasound intensities, you break the shell and this perfluorocarbon gas comes out and you now have free gas bubbles.

If you apply more ultrasound to this system, these things oscillate. They go into nonlinear oscillations but basically we consider this stable cavitation if the intensity is not too high, but if the intensity is higher, then this bubble can grow into inertial cavitation, get really big, collapse, and now you have the potential for bioeffects, maybe even bioeffects down here at the stable cavitation threshold, but certainly bioeffects here at the inertial cavitation threshold.

[Slide 38]

Here are some of the effects of contrast agents. They are on the order of between 1 and 7 microns in diameter. This is a system where you use a diagnostic ultrasound system to scan. You put the things in a little chamber. This shows you what the effect of contrast agents is at different levels of concentration with a commercial diagnostic ultrasound system.

This is a lite chamber and you see that you have enhanced the contrast. At this particular concentration you have significant contrast enhancement, but look what happens if you continue to increase the contrast. You then get shadowy, because you can have the concentration too high and you are trying to propagate through a bubble cloud, so to speak.

Here is a little tube now and there is a diagnostic ultrasound system that is irradiating these things down like that. One of the things that happens is that you break up these bubbles. As soon as you break up the bubbles and the gas is released, the gas is probably dissolved, these things are cleared, so you see this is nice contrast but, after you apply ultrasound for a little while, it is cleared and the contrast is no longer there, so you are breaking up the bubbles.

You can envision that the first person who invited this -- in fact, it was a guy by the name of Weinstein, I think, at the University of Chicago -- was a physician. What he thought was happening was that the ultrasound went in, hit the bubbles, scattered back. He got lots of scatter and that is how you got the contrast.

It is a lot more complicated than that. For example, one of the things that happens is you break these bubbles and, once you break those bubbles, all kinds of things can happen, so Wen is trying to understand that.

[Slide 39]

Let's just look at this portion right here. This is a case where he is applying ultrasound to a vial of contrast agent bubbles and then using this passive cavitation detector to look at the sort of acoustic emissions from that.

This is part scattering and part emissions, it is hard to detect exactly what is what, but as you increase the amplitude in this direction, you get no acoustic emissions, no scattering, or very little, and then there is a region where you probably break the shell, release the gas. You have an effect. This is what we call a P1 or a shell-breaking fragmentation threshold. Then everything gets quiet again, or at least quieter. This, by the way, is the FFT of that.

If you continue to increase the amplitude, pretty soon you get what we would call sustained inertial cavitation. This is to be avoided if it is in a little baby's veins. This is a potentially high bioeffect. This is not what you want to use if you are trying to do imaging.

[Slide 40]

Wen is trying to understand what happens now as a function of amplitude and as a function of pulse width. One of the things I wanted to tell you about lithotripsy is that once you have cavitation, you have a growing bubble, you have got perfusion of gas into that bubble, and now you have debris, and that is gas-bubble debris, daughter bubbles, if you wish. They are bubbles resulting from the previous event. Those things hang around and they can be the nucleation sites for the next acoustic cavitation event.

If you have a short pulse length, then nothing much happens in here when you break up the shell. If you have a longer pulse length, then you can start to see some behavior in here that, after you do something with them, it goes way. You see that this sort of falls down after a while.

The idea is that you break the shell, the bubbles come out, you cause them to become activated a little bit, you break them up into little pieces, and they sort of go away. They do not go completely away but they sort of go away.

Now, when the amplitude is up very high, all those little bubbles that are broken up from the previous event, now the amplitude is high enough, the intensity is high enough, the threshold is slow enough, they just sit there and cavitate and that is a dangerous position to be in.

[Slide 41]

If you have high pressure amplitude, which is this curve right here, you can have sustained, continuous cavitation that can cause damage. If you have lower pressure amplitudes, it can come up and then it goes away, you use up the bubbles, or if you have a long pulse length

-- at short pulse lengths the bubbles break up and sort of become inactive -- but if you have a long pulse length, you continue to activate the bubbles.

Anyway, this is work in progress from Wen and he is still coming along on his dissertation.

[Slide 42]

Here is some work from Sandy Poliachik. Again, I am not doing justice to some of this stuff but I wanted to show you some of the things that they are doing. What Sandy is interested in is whether you can perform what we would call acoustic coagulation. In the next lecture I am going to tell you about how we use ultrasound to stop bleeding.

What Sandy has discovered is that if you apply high-intensity focused ultrasound to blood with various levels of hematocrit (the amount of blood cells in the dissolved fluid), you can activate blood platelets. Platelets are the things that start the initial coagulation scheme to get a clot to stop you from bleeding.

Here is platelet aggregation; that is to say, these platelets are sticking together, one of the first stages in the formation of a clot. This is the threshold, which is a pretty high threshold, but without contrast agents, and she starts to get aggregation as she increases the pressure amplitude here, or some sort of intensity. It goes up like that. As soon as you add contrast agents, you back this up and it goes much, much faster.

She has tried to correlate this in terms of what we call the cavitation dose so that we can do some sort of acoustic emission system, determining how much energy you can apply to the system and how much feedback you get to give some sort of dose response.

Here is a kind of intriguing figure. These are little clumps of blood platelets and these little strands in here are strands of collagen. If you cut yourself, the blood platelets get activated, they suddenly recognize they are in a high flow field, so they activate and they send out little sensors. Then they all join arms, they aggregate, and that happens a lot in your bodies. You get platelet aggregation all the time as you go through the heart valves, but you do not want to have clots running around in your bodies.

The way they actually form a clot is they see exposed collagen. If they see exposed collagen, they rapidly grab onto the exposed collagen and that is what you want to have happen if you cut yourself.

DR. GARRETT: To quantify the platelet aggregation, why don't you shoot heparin in there and titrate it to determine what concentration of heparin nullifies the ultrasonic coagulation?

MS. POLIACHIK: It is an anticoagulant. We actually use anticoagulant in the blood samples that we take and then this is overcoming that.

DR. CRUM: We are bypassing that actual stage. What does heparin do? Does it jump over the von Willebrand's factor?

MS. POLIACHIK: It stops one of the proteins in the platelets.

DR. CRUM: Maybe I will show you later that we bypass a lot of the stages in coagulation when we use ultrasound. We can do coagulation with or without heparin, it is not a problem at all.

That is a good comment, Steve, in terms of trying to quantify this sort of stuff.

[Slide 43]

Let me move on here and tell you about some other things. Here is something that we want to try to understand and that is how contrast agents work themselves to make a better contrast agent. One of the things that Tom Matula and some of his students are trying to do is to look at an individual contrast agent, because we want to know the acoustic characteristics of those individual contrast agents.

Let's suppose you acoustically levitate it, as Tom does, single-bubble sonoluminescence. You do a levitation of a single contrast agent bubble and you can do this at, say, 20 kHz, and then you have in here a 2 mHz transducer, also, and you ping it. What happens is, if you do light scattering off of it, you can actually see the radius-time curve of a single contrast-agent bubble. This shows you not a single contrast-agent bubble but the capabilities of a light-scattering system -- truncated -- but here are the rebounds of the system.

In this particular case I showed where if you had a breakoff of a microbubble you can actually see that, so the resolution of the light-scattering system is enormously better than any way people have now for looking at single-bubble contrast agent bubbles. We have an NIH RO1 that is going in, in a few weeks, to study this whole problem of the individual bubble dynamics of ultrasound contrast agents.

By the way, there were 17 companies building ultrasound contrast agents and the capitalization of those companies was approaching about a billion dollars and they were all about to go broke because the diagnostic ultrasound companies used the idea of the harmonic effects of bubbles to do tissue harmonic imaging, so they do not really need these contrast-agent bubbles as much as they used to. Nonetheless, this is a huge industry. The whole contrast-agent industry, not ultrasound contrast, is a several-billion-dollars-a-year industry and we are trying to understand something about that.

[Slide 44]

I have a lot of things to cover, and here is something else. That was sort of tip-toeing through some of the graduate students' projects, and now let me tell you about some of the things we would like to do in the future and some of the stuff that we are doing now.

If you do high-intensity focused ultrasound, and I am going to use that as a noun and a verb in all kinds of things, HIFU, if you HIFU something, we have high-intensity focused ultrasound, focusing it down on the tissue here, and instead of doing scanning and imaging, we are actually going to try to do therapy.

If you take this transducer right here and focus it down and the intensity you use is kW/cm^2, not mW, which they use in the diagnostic industry, you actually can cut holes in liver tissue. If you move it around, you have so much absorption at MHz frequencies you cook the tissue. If you can cook the tissue, you can stop bleeding, and I will tell you about stopping bleeding.

[Slide 45]

Let me tell you about the potential of what I call bloodless surgery, or HIFU surgery. This is a nascent industry. There are two or three little companies that are trying to do this sort of stuff, but I think it has fantastic potential.

The idea is that you have a HIFU transducer, a coupling medium. You focus the ultrasound down to a spot. This spot can be controlled, depending on your aperture, your geometrical gain, your phasing of this thing, to the size of a grain of rice, not much bigger than that. You can propagate the ultrasound down and get a focusing of the ultrasound here so that you have a spot something like this.

This could be either the temperature -- this is actually the acoustic field, the intensity, and then you convert that to a bioheat equation and you get a temperature field.

[Slide 46]

Let me tell you a little bit about how you can do this stuff and let me show you an effect now. This happens to be the lens of a cow's eye, so you can see through it. This is about a centimeter in diameter this way, maybe a half-centimeter the other way, so it is a completely transparent medium.

The movie will not turn on. Let's skip it.

[Slide 47]

The idea is that if you apply the ultrasound by what we would call image-guided ultrasound, here is an imaging system, so this is what you see with the imaging system, and suppose that is, say, a metastatic tumor inside the liver itself. Then, when you apply the therapeutic ultrasound, this indicates that you are apply the therapy and the imaging simultaneous (actually, sequentially). You can actually see a spot appearing at the focus of the HIFU that shows in the imaging, so you can guide the therapy to a particular spot.

Here, we want to damage this tissue inside the liver itself. We went through the skin of the pig like this, no damage here -- no damage here -- no damage around here, only at the area where we are trying to apply. Gail Truhar, whom some of you know, at the Institute for Cancer Research in London, has actually treated almost a hundred patients with this sort of stuff now, but without image guidance. Our claim to fame is image guidance.

[Slide 48]

The other thing that you can do in terms of bloodless surgery -- on the front page of one of the newspapers yesterday was that they have a new blood substitute that will allow people to do bloodless surgery, I did not read the thing far enough, but here is how you do bloodless surgery.

This is a lobe of the liver. What you want to do is cut that lobe of liver off, because there is cancer in here. You cauterize right there. We apply ultrasound down like that and we cauterize the tissue with ultrasound. Now, when you cut through there, you can take that lobe of the liver off and it does not bleed.

For a lot of people bloodless surgery is a very attractive thing. When you have a splenectomy, taking the spleen out, you often use 7 units of blood, so being able to do a splenectomy without using 7 units of blood would be a very desirable thing.

[Slide 50]

One of our real challenges, and Mike Bailey, Shiram Voisay, Roy Martin, and a whole bunch of people are working on this, is could you do image-guided therapy and monitor the therapy at the same time? Here is the real challenge and we are working really hard on this.

Here is the HIFU beam. It is producing a lesion, and this is in vivo now, because we are scanning, and this is in the liver of probably a pig and we are actually seeing the spot where the lesion is occurring. If we could monitor this lesion occurring in time and there was a tumor, let's say, right there, we could just sort of paint out the tumor with the HIFU and cause a coagulative necrosis, or destruction of the tissue. [Movie]

We have used various techniques for lesion imaging, so here is the thermal beam itself, and this is all scaled, so that everything is at approximately the right scale. If you do just an ordinary diagnostic ultrasound scan of that, you see this little area right there, that is probably due to bubble evolution outgassing as you heat this thing up to 70°. You have to heat it up to 70° or 80° to kill the tissue.

If you do what is called collar-power angiogram, which looks at the flow of blood due to a harmonic Doppler effect, you see this area right in here. If you do what is called flash-power Doppler, that is to say, any time there is an emission coming out, if the cavitation is sending out a pulse and there is a Doppler transient, then you try to correlate that -- you see that is all uncorrelated here, there is no collar, but there is nothing coming out of the area in there, because all the contrast-agent bubbles are no longer alive there, they are destroyed.

Now you are starting to see a real image of the lesion and if you cut this thing open and look, there is the lesion. This is almost to scale now, so we are very closely getting to the point now where we can actually guide the image to the site of the tumor.

[Slide 51]

There are lots of nice acoustics to do in this and I am going to skip through some of this, but I want to show you some of the nice acoustics that we have done. If you shoot a long time, what happens is that you probably get a cavitation bubble and then you cannot make the lesion go beyond that cavitation.

If you shoot slow and at a lower intensity, you can actually make a nice long, well-controlled lesion. What happens if you have a cavitation bubble that occurs right at the focus? Francesco Curra, a graduate of PASS two years ago, has done lots of analysis of computations of the lesion formation, and I will talk about that a little bit more.

If there is no bubble in there, he calculates this sort of shape, but if you put a bubble in there you get lots of backscatter and you get a tadpole-shaped lesion, so we are starting to model that. There are lots of nonlinear effects here.

[Slides 52 and 53]

If you are depending upon the nonlinear characteristics of the tissue itself, and various tissue types have various nonlinear characteristics, if you have very little nonlinear, the B/A, and all these sorts of things are quite low, like blood, then you do not have much of a nonlinear effect and the sound propagating from the HIFU -- you look at the various harmonics here and they look like this -- look what happens if you add more and more nonlinearity to the tissue. You get this very dramatic shock wave. The dramatic shock wave has lots of higher harmonics in it.

The tissue attenuation depends very strongly upon the frequency, so you can control the amount of absorption at the focus by controlling the amount of harmonic content.

DR. ATCHLEY: What is the z/F?

DR. CRUM: That is the distance between the focus and the point it is looking at, so z/F of less than one is upstream, z/F of greater than one is downstream, and z/F equal to one is right at the focus. [Movie]

This is a movie of the propagation through three layers, water, fat, muscle, and liver.

[Slide 54]

He has done this now in three dimensions. This is, I think pretty sophisticated nonlinear propagation. Vera Klokova and Oleg Shapoznikoff, who come out of Rudenko's group, have taught Franco to do some of this stuff. He starts with a transducer over here. He has built this so that he can take into account an array, so he does not have any kind of geometry required. He allows both forward and backward propagation, so he is not doing the parabolic equation or the KZK approximation.

He allows an attenuation coefficient to have a power law, because some of these different tissues have power laws in terms of attenuation of 1.2, some have 1.4, some have 1.1, and he also then couples the bioheat equation to the acoustic propagation equation and so these things down here are actually calculations of lesion size, depending on whether you use a temperature criterion of 75° or a thermal dose criterion of 120 minutes at 43°; that is the thermal dose criterion.

Now what we are trying to do, then, is model the propagation of these waveforms inside and calculate the lesion itself.

[Slide 55]

I want to show you the sorts of things that we are trying to do in terms of actually building systems to do that. In terms of commercial systems, what we want to do is to try to make transducers that actually can produce some of this type of HIFU effects.

One of the things that you can envision (I am going to show you this later) if you wanted to stop bleeding in a particular area, you would want to do that with a device that even a doctor could use and could not damage. Now we have actually built a system where this is a holder, this is the transducer, and this is a waveguide, an aluminum or brass waveguide. This is a very durable system and this is the kind of waveform we get with that, the sort of beam profile.

You get a very strong beam here that is relatively focused in this particular direction, and here is the profile of the intensity pattern. This is removed from the tip. Depending upon the shape -- actually, you can make this concave or convex -- you can shape the tip of this beam profile so that you can apply it in some sort of what we call an intraoperative acoustic hemostasis technique.

[Slide 56]

One of the biggest industries that could develop is catheter-wound sealing. Five million people a year get a catheter run up their femoral arteries into various regions of their bodies, particularly their hearts. If you have a stenosis in one of your coronary arteries, the first thing they do is they go up and look at it, or they do some sort of angiogram.

[Slide 57]

They run that catheter up in there and they do a balloon angioplasty or they just look at it, so they do lots of catheterizations a year, 5 million catheterizations a year, run the catheter in the femoral artery up to the heart. When they take that catheter out, you have got a hole in your femoral artery, which is a strong artery, and if you allow that to bleed, you will die. How do you stop the bleeding, how do you induce coagulation in that particular femoral artery when you pull the catheter out?

The way they do it now, the state of the art, is put your thumb on it (you actually put three fingers on it, one finger upstream to stop the flow, another finger down to hold it like that, and then a finger downstream) and hold it for a little while and then put a sand bag on it with some sort of strap around it.

Quite often, if you talk to someone like my father-in-law, who had a catheterization, he said, "I had to lie there for eight hours with a sand bag strapped to my leg." That is what he remembers from the balloon angioplasty.

There is a whole industry for stopping that bleeding. One company, called Perclose, actually put a collagen plug in there. You run it down in there and you stick that plug in there, it has little fish hooks in there to hold it. It is a very invasive scheme to try to do that.

With this company we are working with now, Ferris, we think we can do it this way. We could focus the ultrasound in both an imaging and therapeutic way and when that catheter comes out, we fire the ultrasound down and we just cook that tissue right there, and that seals it. I will talk about how we seal things with catheter wounds and acoustic hemostasis in the next talk.

We have been able to generate, or at least they have told us they are going to give us $4.8 million from the Army to try to stop not this problem, but the acute arterial wound bleeding in the combat casualty care situation, where we are trying to solve this same problem. This is the sort of approach that we are trying to use.

I want to take just a few minutes for discussion and then I want to get into the specifics of the particular case of acoustic hemostasis.

I want to summarize this whole thing by saying that I think there is a tremendous opportunity for physical acousticians to go into the therapeutic ultrasound industry. Diagnostic ultrasound imaging, when you see things and you can make a diagnosis, that is great, and that is a $3-billion-per-year industry and it is getting bigger and better all the time.

If you could use this capability of ultrasound to go in, with damage to intervening tissues, and have a biological effect, a controlled biological effect, at depth, the whole idea of noninvasive image-guided transcutaneous bloodless surgery is all in front of us. This is going to be done. Probably I will not be involved in a lot of it, but some of you could be, and I think there are tremendous opportunities here and we should follow through on some of these opportunities and go for this whole new industry that is going to develop in the future. A lot of good physical acoustics has to be understood before we know how this is going to work. It might work without it, but it would certainly work a lot better if we understood the physical acoustics.

Comments? Questions?

MS. SWEARINGEN: When you are talking about using contrast agents for some situations, how small does something need to be before they need to use a contrast agent or what areas of the body are they more used in?

DR. CRUM: You try to make them about the same size as a red-blood cell so that they will go through the body and will not block them, so you have to keep them about less than 7 microns in size.

One of the attractions of this thing is that you can do venous injections. If you do venous injections, it goes in through here, it goes to the heart, goes through the lungs, and then goes to another particular area -- well, it goes into the upper portion of the heart, then it goes to the lungs, and goes back into the ventricle. Then it goes to the coronary arteries, so it would be nice to be able to do venous injections.

Of course, you can put it into the carotid artery o the various arteries that go directly to the left side of the heart, but it would be very nice to have these things that stay in the body for a long amount of time.

Contrast agents are now being designed to go to particular areas. For example, 50% of human males over 50 have some form of prostate cancer. It might be just hyperplasia and preformation, but humans are not meant to live beyond 50, so they develop these cancers.

If you tag -- you have heard of prostate-specific antigen, PSA -- PSA is a tag that says I see a certain specific antigen inside the prostate itself. You can tag a contrast agent with a prostate-specific antigen on the outside, it goes to that particular area, and now you do an ultrasound scan and the area where you might have cancer inside the prostate just lights up, so you now can do very specific imaging of that particular site.

Now think about the next thing. Suppose you took that bubble that was starting to stick to the cancer cells in the prostate, you could image it there. Just hit it a little bit harder with HIFU, then you are going to start killing that tissue. Or if you do not want to get so active like that, just have a prostate-specific chemotherapy drug that gets released only there. That is what we would call site-specific drug delivery in a particular area.

All of these contrast agents provide ways now of carrying drugs. Rather than using them as contrast agents for ultrasound, they can be site-specific drug-delivery vehicles.

MS. SWEARINGEN: I guess the question I was asking was how big, typically, is an object that you are imaging? How big does it have to be before you do not really need the contrast agents to see it well?

DR. CRUM: I do not know how to answer that question in the sense that the resolution that one has with diagnostic ultrasound systems depends on their frequency. If you go to transvaginal imaging with 7.5 to 10 MHz, you can see eggs moving down the Fallopian tubes. If you want to try to do just images of a baby, you use 2.5 MHz and you can see the nose and ears, so you are talking about various dimensions, depending upon the frequency, resolution, and so forth.

I did not mean to be a politician there and just go on and talk about other things; I really did mean to answer your question.

Time for a break.

DR. CRUM: This work was done by Lisa Couret and Cyril Lafont and I think it is a very interesting movie, because you can now see, in a sense, what HIFU does in tissue. [Movie]

Over here is a transducer. This is a lens, maybe a centimeter in diameter, maybe 8 mm, or something like that, 4 or 5 mm in depth. These are little holders. Remember that sort of focus is in here somewhere, it was hard to tell where the focus really was.

This is like the albumin in an egg. You can almost see through an egg, the white of the egg. Obviously, the lens of the eye you can see through, but as soon as you heat it up to about 65° or 70°, these proteins cross-link, in the same way as when you cook an egg, it gets stiff and is no longer opaque. What we are doing here is cooking the proteins, cross-linking the proteins, in the lens of the eye. Now you can see quite readily this lesion forming.

What Cyril has done is calculated, in a sense, the size of this thing, the intensity it requires to start it, and the expansion, so to speak, in time. You can see that it has an evolution in the sense that it will start right here, moves downstream, stops there, backs up a little bit, now expands in this direction. You can see all the good thermodynamics and acoustics occurring there, and all in one simple little device, a nice model, we think, for studying this therapy stuff.

Now I want to go back and do a specific aspect of this HIFU therapy. This is the case now of this battlefield combat casualty care. We were very successful in building a company, even, to make these handheld diagnostic ultrasound devices. They are distributing these things now in Kosovo and in Korea and on Navy ships.

One of our undergraduate students just got a NASA KC135 flight to carry one of these things in zero G, because what happens in zero G is you no longer have gravity pulling your organs down, to try to measure what is happening to your organs as you go to zero G in terms of the redistribution of these things.

Now, could we make a system that could solve the problem that one has had since they started killing people with sharp objects; that is, if you cut one of your arteries and it is a relatively major artery, unless you can stop the flow, the coagulation, the blood-clotting sequence, is not going to stop you from bleeding to death in many situations. In particular, if it is an aorta or a femoral artery, or some of the major arteries inside the abdominal area, you are going to bleed to death.

Since the Crimean War, when Florence Nightingale first started doing statistics, about 40% of the people who die on the battlefield die within the first 10 to 15 minutes from exsanguination, which means bleeding to death, so that is a problem.

We proposed, whether or not we ever get a solution to this problem, that we build, first of all, a small device for imaging those systems like this (you can now see it), and then HIFU-ing it and stopping that bleeding, so now we want to look at this stopping-the-bleeding problem.

[Slide 58]

Here is the problem in one slide. If you lose 2L of blood, you go into cardiovascular shock -- that is this line right here. You have about 7L in your body, so 2/7ths of it is enough to kill you if you lose it. If you cut a major artery, like a bullfighter's wound when the bull rips out the femoral artery, you are bleeding at 400 mL/min, you have less than 10 minutes. That is why a lot of bullfighters die when they get gored in the groin.

Now, if you run into a tree and run a post up your chest, then you might get into something like this, which is on the order of 40 mL/min, and now you have an hour to stop that bleeding, so these are various levels around here.

Our idea was that if we could slow the bleeding rate down into this area here, then you would have an opportunity for what they now call FAS teams, forward Army surgical teams. There is a nice Army expression for it, "grab and run," or something like that. The medic is supposed to go in, grab the person by the collar, pull him back, put him in a Humvee or some sort of armored personnel vehicle, and then you would try to stop the bleeding right there.

If you could do something like that to stop the bleeding, where you could evacuate them to an Army MASH unit, where you could do surgery, you could save these people's lives.

[Slide 59]

I showed you earlier the capabilities of diagnostic ultrasound for imaging the vascular system, and here are specific ones. This is called power angiogram, which means that you collar anything that is moving, so here is blood being irradiated into the entire kidney itself. You can see all of the perfused area.

If you go to 3D, you scan and then you move the transducer over and you scan again, store all this, and reconstruct the image. This is a 3D collar-power angiogram of the blood vessels inside the liver.

These are the blood vessels inside the kidney. You can image the entire vascular tree and see vessels down to the order of a few millimeters in diameter. If you had enough time, and you do not, you could actually see a lot of these bleeding sites.

[Slide 60]

Rather than do all that, another way of doing it in terms of solving the exsanguination problem is, first of all, determine how much someone is bleeding. The bleeding rate is extremely important, so that is the first thing that you want to try to do.

They are actually using this system now in the Level I trauma center at Harborview Hospital, which is a Level I trauma care unit in the University of Washington medical system. It turns out, on a Friday night in Seattle, you still get two or three people who are shot or who run their cars into trees and are bleeding internally, and you want to know how much they are bleeding internally.

This is being used, in fact, all over the world. I just came back from Florence and people in Florence were using this system to do the following: This white area here is the amount of fluid that is gathering in -- I think this is the Morison's pounch, which is an area underneath your liver here -- yes, Morison's pouch, so you can measure how much blood is being filled into that area.

If you can measure the volume of that as a function of time, you can say, well, the person has an hour before he is going to bleed death, or the person has 10 minutes, let's do something, go Code Blue, or whatever they do in the E.R., and go in and do that.

[Slide 61]

In a military situation, first of all, you have to do this triage, or diagnosis. The second thing is, now if you see that someone is bleeding very rapidly, to try to find the site of bleeding. Again, ultrasound can do that. This is a pseudo-aneurysm. A lot of people have pseudo-aneurysms, your aorta has a little weak spot in it and the blood is squirting out and coming back, squirting out and coming back. If that breaks into a full aneurysm and you start bleeding internally from the aorta, you are going to die and there is not much you can do about it.

One of the things they are doing at the University of Washington in the vascular surgery department now is just random screening of people. They go in and they scan these major arteries to see if you have a pseudo-aneurysm. Here Kirk Beach and some of the guys doing this have detected one, and this is an artery here and there is a little weak spot in there, so the blood is squirting out and coming back.

If you do Doppler and you collar something going up, up-Doppler is red and down-Doppler is blue, you can see turbulence and you can see, then, a pseudo-aneurysm, but the point is that Doppler not only tells you that you have this bleeding site but gives you spectral Doppler and you can tell the volume, so you can tell magnitudes, and so forth.

[Slide 62]

One of the things that this guy, Kirk Beach, discovered was a nice structural acoustics-type thing. Imagine that you have a pipe and you have a leak in the pipe and the water is squirting out of the pipe, you can go in and measure, with a Doppler system, the blood coming out. You have movement there and you can look at the Doppler effect and it is all very nice, and that is what they do.

What he did was go in and look at the pipe itself and determined, as the water goes out, that you get eddies formed and the pipe vibrates. What he saw in there was low-frequency vibrations on the order of 200 Hz, which are the structural vibrations of the artery itself as a jet of fluid comes out.

Those low-frequency vibrations are bruits. Some people who are really good -- Kirk is an M.D.-Ph.D., and he said, "You know, in the good old days people could hear these with stethoscopes," because they actually make enough noise, you can hear them. We think that this might be a way of going in and determining whether a person is bleeding at a particular rate or not.

[Slide 63]

You see here is a very complicated figure that Kirk has made in terms of explaining to the military the various ways that he is going to use to determine the various detection schemes for bleeding. In some areas you would use collar-flow imaging (that is that pseudo-aneurysm Doppler thing that I showed) and down in here is vibration detection, and down in here is expansion.

The other thing that you can determine, for example, if you are bleeding in the brain and it is all encapsulated, that is bad, because you have the skull here and that prevents the blood pressure from being released, so if you get a subcapsular hematoma on the brain, what you do is you apply pressure on the brain that then collapses some arteries going to that portion of the brain and your brain dies.

One of the things you want to do is cut a little hole in there and take that blood out to release that pressure. You look at the expansion of the brain, or the movement of the brain, and Kirk contends that he can detect high-frequency nanometer movements. Anyway, the idea is that you can see a lot of these bleeding sites with ultrasound.

[Slide 64]

I have already talked to you about contrast agents and how you can see bleeding with contrast agents. This is another picture of some of our experiments in which we are using contrast agents to detect bleeding.

[Slide 65]

I want to show you now how we can stop bleeding, how we go about stopping bleeding. I mentioned to you before that if you have this focused ultrasound down and you leave it at one spot, you get cavitation and you get debridement and you cut a hole in it, but if you move it around, you cook the tissue. If you cook the tissue, you can, in a sense, do the old John-Wayne-movie type thing, he gets shot and he is bleeding, so John Wayne sticks the poker into the fire, gets it red hot, and jams it into the wound. That is cauterization. It worked in the movies.

[Slide 66]

Here, now, is the blood-and-guts portion of this particular sequence. I am going to show you how we can do what we call acoustic hemostasis under a variety of conditions.

This is going to be the old stuff in which we stop bleeding in the liver.

This is going to be punctured blood vessels.

This is going to be lacerated blood vessels (and I do not think I am going to do the spleen).

[Slide 67]

Roy Martin, one of our guys, always wants to say how successful this stuff is and a big thing in the biomedical community is your N value, because it costs you $500 to do a single animal, a large animal, so it is expensive to do lots of animals. If your N value is 1 or 2, it means you do not have very much money, you cannot do very many animals, and nobody believes you.

If your N value is 15 or 20, this is a lot of pigs he has killed, and that is not good, either, because the animal rights people do not like your N values to be up. The point is that in order to get the scientific community to believe you, you have to have high N values and a lot of these different statistics.

I put that in just to point out to you that our N's are very large for some of these; we have killed a lot of pigs.

[Slide 68]

Here, now, is the first sequence. This is a bunny rabbit, a white rabbit. By the way, everybody should be sensitive to this. At the University of Washington, which has one of the largest funded biomedical programs in the country (I think it is No. 2 or No. 3 from NIH in terms of funding), $350 million a year to the medical school, at every animal experiment that is done you have a veterinarian or an animal rights/animal care person who sits there and holds the rabbit.

In this particular case, this woman would stroke the ears of the rabbits and when they sacrificed the animals it is done as well as you could in terms of protecting the rights of these animals, or at least to do this in the proper way, so I do not mean to show any lack of concern at all for these animals. We are trying to do these experiments in the most humane way we can.

What we are doing here is this rabbit's liver is exposed, it is under water. Here is a transducer and now we are going to cut the liver.

[Movie]

This rabbit, because it is bleeding under water, would bleed to death. Now what we have done is we have a transducer that is, again, under water, and it is focusing the ultrasound at pretty high intensities now -- we are talking about kW/cm^2, so we are talking about 50 or 60, maybe even 100, atm of pressure at 4 MHz. That is just cooking the tissue.

This would be a tremendously difficult wound to seal, to stop bleeding. You might say, well, this is just a rabbit and why would you do something like that? There is liver surgery all the time and the people we have involved in these programs are liver surgeons who try to solve the problem of how do you stop bleeding in the liver.

Can you envision trying to sew the thing up? You cannot sew it up, you cannot suture it. It is very hard to get to the bottom of these systems and apply these bovies, which are high-frequency microwave systems that cauterize the surface, so getting way down to the crack at the bottom here is almost impossible.

The way they do this is they do some cauterization on the surface, then they stuff it and they pack it. If someone is doing liver surgery on you, very likely he is going to stuff your liver, pack it, with all kinds of tissue and let it sit there for 40 minutes to an hour. Then they go back in and very delicately pull that out and see if your body has coagulated and stopped all that bleeding. That is the only way they can do that.

Now look what we can do here. We can just cook the tissue with this ultrasound. What I have done here is cut out about 30 seconds and in about a minute we were able to cauterize this wound along here and we were able to keep that rabbit in the water for as long as humanely possible, which was about an hour, and it did not start bleeding, again. That is the first case.

This is gross. This is a liver of a pig. This is what happens if you slam your car into a tree and your steering wheel hits your liver. It is going to fracture the heck of it. This liver is all fractured here. There is a technique for actually doing this involving a metal plate and a nail gun for this pig -- that is how you do it (transcutaneously, believe it or not).

[Movie]

What we are doing with this transducer is we are moving it around on top of that area there and we are actually cauterizing the tissue at depth and all through the surface. If you apply that ultrasound there, it just cooks the tissue. You might say, gosh, that is damaging the tissue. Well, if you do not have a liver, you are going to die, so if you could damage part of it and stop the bleeding -- oh, gee, I missed that, I did not think it was coming up that fast.

What we did in the liver is we cauterized one side of the liver and it essentially stopped bleeding and the other side we just let bleed and it bled for as long as we could humanely keep the animal open. It never stopped bleeding. We were able to demonstrate in that case that we were able to stop bleeding in a fractured liver.

Now we cut a hole in this femoral artery and the blood was squirting out. We are applying the ultrasound now through this very crude device, moving it back and forth trying to stop this bleed in a femoral artery. The femoral artery is as big as my middle finger, or bigger. The cut we made was 7 mm, almost a centimeter, in diameter, and this blood squirted out in an arc of about three feet. This pig would have died in about 10 minutes.

[Movie]

I want to show you how fast. This is real time now. We are shaking this transducer around to get maximum coverage and you can see that spot right there. Watch very closely after he takes that off. There, right there is that sealed point, we have sealed it, and you can see the heart beating.

I did not run the volume up on this and you are not going to be able to hear this, I think, unless you listen very carefully.

DR. COSTLEY: What frequency is that?

DR. CRUM: Somewhere between 2 and 5 MHz. We have different transducers, depending upon how deep, but anywhere between 2.5 and maybe 7 MHz, in some cases, but these are all probably 3.5. [Movie]

Now here, I thought, is an ingenious thing that Roy Martin did. If you listen carefully, maybe you can hear it. The idea is this: They punctured the artery, the blood is squirting out. You apply the transducer down. This is about the size of a small catheterization.

When you apply the transducer down on top of the thing, where is it bleeding? You cannot see, so you move it around. If you move it around, that spot is about the size of the hole here, so you are cooking all of the tissue around there. How do you aim a little bit better?

In this transducer they are applying they use the system as a Doppler system; that is, they send out a pulse from the therapy transducer, listen back for the frequency components, and you can hear this, you can play this. If you listen very carefully, maybe you can hear it. You can hear the blood coming out toward the transducer.

Every time the heart pumps, it goes "cch, cch, cch." If you move this thing around until you maximize the amplitude, then you know that the therapy system, which is focusing down, is right on top of the hole, and then you go "uuum," and you are going to hear that, and that is firing the therapy system trying to hit the spot, so maybe you will hear all this. [The audio segment is played.]

That worked. He moved it around and all it took was about one second of a HIFU blast and we stopped the bleeding. That is why we have some confidence, or we can at least argue to the military, that we do not need a lot of power here.

DR. WILEN: In a situation like this, are you coagulating the blood in the vessel or are you cooking the tissue of the vessel wall?

DR. CRUM: I will show you. You did not know that was coming up unless you were looking ahead.

[Slide 69]

Here is a blow-up now of the artery and here is that wound -- right there. When we pull the needle out, then tissue comes out and probably sloughs back in there. When we apply HIFU, we cook this tissue here. Why don't we cook the inside, why don't we cause occlusion inside the blood vessel? That is what we were worried about.

If you stop the thing from bleeding and you shut of all the blood flow to the leg, maybe it would be better to bleed to death. What happens is that we cook the tissue right in the hole, so this is necrosed fibrin tissue up in here, but if you look in really close, we find that in the interior this endothelial tissue here is still alive.

We have not done survival studies. One of the most difficult things to do when you do large animal studies is survival studies, because you can imagine stripping open this pig and going in and getting these arteries and punching holes and sealing them up, then sealing the whole pig back up, again, because you have to do this all sterile, then let the pig run around; the animal care people find it very difficult to allow you to do survival studies on something like this.

In this particular case you can leave the animal exposed for several hours and there is no rebleeding and it looks as if this tissue is still alive there. Why is that true? Physics proves it. One of the nice things about having physicists working in this particular area is that we can do quantitative calculations.

[Slide 70]

Francesco Curra, one of our graduate students, did this calculation. Here is the transducer -- over here. If you are up here close enough, you can see some of the field coming in. He has converted this to a temperature field, so it is warming up a little bit here. Without a blood vessel, here is the temperature profile of the focus of the HIFU therapy transducer.

If you put a blood vessel in there, look, there is no high temperature in there at all, it is really cool, 45° to 50°, inside the blood vessel in this dimension. Why is that true? Two things. These Powerpoint slides have their limitations, but if you look up there you can see a stream of blue collar going up in there. Well, the blood is going up in that direction, so two things are happening.

One is, the absorption coefficient of sound in blood is very low compared to the absorption coefficient in tissue, so there is very little sound absorbed in the blood. The second thing is the blood is flowing through here relatively fast and it cools the inside, so two good things are happening there to prevent damage to the inside of the blood vessel.

That gives us some hope that we will be able to do this, actually, on humans and that there will not be any long-lasting deleterious effects.

[Slide 71]

Here are some data, again, to make it scientifically rigorous in terms of numbers. Here is something that Steve Garrett asked. If you do heparinized and non-heparinized animals, the difference between the coagulation times, or, as we call it, the hemostasis times, either complete or major, is independent. The red bar here is heparinized, the blue bar is non-heparinized animals, and there is no difference in terms of the complete hemostasis times, depending on whether they were heparinized or non-heparinized animals.

These are the times and the times required to do this stuff here are on the order of 20 to 30 seconds in this particular case.

Roy and Shiram built this Doppler system so they could listen to where the blood jet was coming out, and when the blood jet was used with Doppler-guided HIFU, 80% of the wounds of the bleeding sites were stopped in maybe 10 seconds -- here is 20 seconds. You see that almost 100% of them were done in 30 seconds, 80% of them were done in about 10 seconds, either with heparinized or non-heparinized systems.

[Slide 72]

The really difficult problem that was thrown to us was, okay, you so have a bleeding site in a blood vessel that you can see and you can apply HIFU. Well, I can do this with hot pokers. Why is ultrasound of such potential benefit? The potential benefit of ultrasound, of course, is that you can do it transcutaneously, you can focus it, you can do all the imaging, and everything like that.

Here is our real challenge and I am going to show you a movie of how we were actually able to do this, a demonstration of image-guided transcutaneous acoustic hemostasis.

[Movie]

We went in and punctured an artery, the femoral artery, on a pig -- here is the pig. We pulled the catheter out. We used this probe here to image the bleeding site. There, when we pulled the probe out, you can see the bleeding. We used this therapy system here to apply therapy -- there you can see the spot.

We dragged this spot on top of the bleeding site and sealed the artery in vivo. Afterward you can monitor with a Doppler and see that there is no blood coming out here and you can take everything off and show that there is no blood coming out here.

[Movie]

This movie has sequences of this thing and it shows different aspects of the whole thing. Here we are pulling the catheter out. This is the imaging probe now. Even by applying pressure this thing is still bleeding, because that was a pretty big catheter, a 14 French, or something like that. You can see it is still bleeding.

Now, when we apply the Doppler -- this is called duplex, this little pie-shaped thing here is the Doppler, the rest of this stuff is diagnostic ultrasound, so you do these things together, so you are imaging and doing Doppler at the same time. As you pull this catheter out along this path here you are seeing the blood squirt out and you can watch this, you can even see the heart beat. See, you can see it pumping there.

The next thing we are going to do is apply HIFU. Now the HIFU is going to come in and you are going to see a little spot develop. When you have this little spot develop, that tells you where the focus is. You take that focus and you drag it over to the spot that is bleeding.

Now, when you turn the HIFU on, you see this outgassing, we think, of the HIFU spot, and you drag that over on top of here, fire it there -- I have cut out some of this stuff -- and you see there is no longer bleeding. These sequences are too long, I just wanted to show you the end result.

We have done this now in many animals and it was essentially almost every time. We have had just an enormous success rate on this -- those guys have, anyhow.

[Slide 73]

Let me show you one other aspect of this thing that I think is potentially very useful. First of all, we have to miniaturize all this. We are trying to sell this to the military now, so we are talking about a particular situation, that is, combat casualty care.

This is out in some combat zone, so first of all, we have to do the miniaturization. This is, I thought, was really a success or an unsuccess story on our part, being able to build one of these SonoSite units. Using the SonoSite unit, we can use that to drive the imaging system. We cannot use the SonoSite unit to drive the therapy system, but you can envision, can't you, that this could all be packaged into one unit, putting the imaging and the therapy into one unit.

What I wanted to show you here -- first of all, this was a sales thing to the military to say we could use the SonoSite thing -- the other thing is, I can monitor the surgery, monitor bloodless surgery. [Movie]

First of all, we wanted to show you this SonoSite system. You can see the HIFU occurring. We turned it on and do you see that spot right there? That spot right there, depending on what kind of coupling you have, if you have water coupling you have all kinds of cavitation in the water and that gives you a lot of backscatter. You really need to know what you are looking at.

Now this is the liver of a pig. What we are going to do is focus the HIFU right there. This is still a very interesting experimental problem. If you are applying HIFU so that the temperature is rising up to 70°, why do you get all this contrast, hyperechogenicity?

I think that what you are doing is the local value of the dissolved gas concentration is exceeding the saturation concentration, you are actually getting outgassing of gas into little bubbles in that very hot spot, so that is what is giving you the contrast imaging.

Other people think that it is a change in the scattering characteristics of the heated tissue but, nonetheless, watch what happens now when we apply that HIFU. You can just cut out an area. You turn it on and you can watch it developing. Turn it on and move it downstream a little bit and you can just cut through like that.

We are cutting lines through liver. I took that slide out of sequence, but do you remember I showed you that liver where we had this line through the liver? What we are able to do was actually do a cauterization through the liver. Then you can go back and cut that piece of the liver off and it does not bleed.

As one doctor said, "Why would you even cut it off, just leave it that way, it's going to die and get absorbed inside the tissue." Right. So you can see all kinds of things we can do with this.

[Slide 74]

Here is the real challenge. This is going to be not physics but engineering; that is, can you put an imaging array together with a therapy array so that you would have one small package of stuff with which you could imaging (this pie-shaped thing here) and then therapy at the same time and monitor it in real time? That is the real challenge.

We proposed that to the military and they said no way, it costs too much, it was a very expensive program. They said, "Could you just give us something like a pop can that you could carry, give it to a corpsman and that particular corpsman could stop bleeding of acute arterial wounds on the battlefield?"

There is a book that is now the bible, apparently, of the combat casualty care community, called Black Hawk Down. You might remember in Somalia, not so long ago, four or five years ago, they got into a huge fire fight and 18 U.S. Marines were killed, and there are stories about what happened to those Marines.

They were in this urban area and they had all these Black Hawk helicopters trying to get them out of there and they could not, because there were all of these Ugandans shooting at the helicopters and this one guy had his femoral artery torn. Even though they were communicating back and forth with doctors, they could not stop that bleeding and the frustration of that young corpsman trying to save his buddy from bleeding to death is now an apocryphal story about what you need to do to save even one life -- you know, future wars, zero casualty wars.

[Slide 75]

We proposed this, but we do not know whether it is going to be funded or not, but we proposed to do the following, that if you had just a pop-can-type thing with a battery and it would have an imaging system that would scan in a Doppler sort of way to see if you could see a jet of blood, if the jet of blood is coming up, then you have to move this thing around until the jet is coming straight up into the transducer and then you would be listening to that with a Doppler thing and then you would fire one second of burst to try to stop that bleeding. Anyway, that is what we proposed and, hopefully, we will get funded to do that.

[Slide 76]

What we are trying to do, also, is, in working with a company called Ferris, to do this commercial problem of catheter-wound closing. Five million catheterizations are done a year, so we are trying to build this system here to do the commercial problem of sealing catheter wounds.

[Slide 77]

I think I showed you this slide before from the Ferris Corporation in which we are trying to build these systems and we have submitted a whole bunch of SBIRs and dual-use science and technology projects to try to do that sort of thing.

That sort of summarizes all the things that I wanted to show you. That is the end of the lectures and maybe we could talk a little bit about future perspectives.

When we did this meeting in Seattle a few years ago, since I was in charge of the meeting, I also arranged to have a lot of the P.R. people come look at all our research results. At any rate, we got in the news and it was not necessarily a good thing, because I think I received anywhere between 25 and 50 letters from people who had relatives or they themselves have incurable diseases, liver cancer, brain tumors, and so forth.

I got these just anguishing letters. I have two of them posted to my bulletin board to motivate me, one of them from a women who said that her husband, whom she has been married to for 60 years, has inoperable liver cancer and she does not think she can live if he dies. Her only hope is for me to come in and do something, "Would you please, please do something?" When you get letters like that, that is a hell of a motivation.

Let me tell you another story. I do not mean to go on about this, but these are tremendous motivations for some of us. I got this voice mail message and the voice mail message says, "Hi, my name is John Doe and I'm an attorney in Kansas City and please call me immediately." What do you do when you get a call like that? You do not call the guy immediately, you know. (laughter)

For three or four days I would come in and I would get this message. I thought, geez, what have I done, what has my wife done, what has Tyrone done, you know? (laughter)

Finally, I am in my office one day and this guy catches me on the phone. "Hi, I'm John Doe, how come you don't return my phone calls?" I said, "Oh, you're the attorney from Kansas City." "Yes," he says, "I've been trying to get hold of you. I have a son, my only son, he is 12 years old. He has a brain tumor and the doctors say there is nothing they can do about it. I can't go on without my son, he's the only thing I have. I can't let him die."

I said, "I can't do anything for you." He says, "Yes, but I know the problem. You can't use this stuff yet, can you?" I said, "That's right, I can't use it." He said, "It's the goddamned FDA. They're preventing you from doing it, aren't they? I will sue the sons of bitches. I've got to save my son's life." That was a phone call.

It would really be nice to be able to do things like that. I would like to be able to do that, save people's lives.

DR. MARSTON: On one of your slides you showed that on a HIFU you could use a .... Is that standard, or what technology do you use to generate high intensities?

DR. CRUM: We have a guy by the name of George Kileman who has a company called Sonic Concepts who builds our transducers for us. We use different kinds of transducers and he is the transducer expert and he builds these things. The biggest problem that we have is trying to get -- our problem in the future, right now we do not have a problem with just making a curved array so that we get geometrical focusing. If you machine them to a high-enough tolerance, you will get very good focusing.

They are single-focus systems, so it is one spot and you have got this coupling problem, so it is a really difficult problem. You can imagine taking this magnifying glass and you are not allowed to move it very much in cooking that spot, so the ideal thing, as you all know, is to go to an array.

Building arrays that can do both imaging and therapy is a major technological problem. I went to ATL, which makes all these transducers, and said, "You guys are the transducer experts, you make all these transducer arrays, can you make an imaging and a therapy array?" John -- what is his name? -- said, "Of course, we can, but we're not going to do it for you, because this is very sophisticated technology and we are not, probably, going to do that."

Ferris, this company that we have hired, has hired some people away from ATL and they think now that they actually can use some of these composite transducers and maybe some of this new single-crystal piezoelectric stuff that has high performance capabilities to do both imaging and therapy with the same transducer.

Maybe you would have to use some transducers, like DCT4, that has high sensitivity, to go for the imaging capabilities, because you have plenty of power, and BZT8 or a single-crystal transducer to go for the therapy, but that is the engineering and it is something we do not know much about, but it is a real challenge.

If we could solve that problem, then we get into the big money production capabilities and it would be a really good thing.

DR. GARRETT: If you have had so much success with implosive lithotripsy tied to your shock spark, why don't you consider something like that with a deformable mirror?

DR. CRUM: There is Biddlinger -- I can never remember where these guys are from, but there are a couple of groups in Europe who are using lithotripsy for all of this work. They do, I do not know if you would call it HIFU, but they do bloodless surgery with lithotriptors. The effect there is all cavitation.

I contend that cavitation, even though it is heresy for me to say this, is an undesirable thing to have in bloodless surgery. Did you hear that? It is an undesirable thing to have, because you cannot control it, while the thermal lesion is much easier to predict, much easier to calculate, much easier to control. I contend that you should use thermal effects rather than the cavitation effects.

There is a whole group of people trying to do it with short pulses, producing cavitation, and they can control things in a different way.

DR. GARRETT: It is easier to deform a mirror than it is to phase array at high powers.

DR. CRUM: Right. In fact, there is a guy in our lab who has this acoustic lens and you can move it. We even thought of doing that. Engineering is something that we have other people doing and I am not interested in more than giving them specs and saying, "Give me this," rather than in actually trying to do the design for the engineering configurations.

We obviously have opportunities in the future, so if some of you graduate students are interested in this sort of stuff and finish your postdoc, give us a call. Thanks very much for your time.

SONOLUMINESCENCE

Thomas J. Matula
University of Washington

[Transparency 1]

DR. MATULA: First of all, I would like to thank Dr. Hargrove, Dr. Bass, Dr. Atchley, the organizing committee for inviting me, and Dr. Libby for helping organize everything.

I am going to talk to you today about sonoluminescence. I have heard from several of you who have never heard of sonoluminescence, so this will be a good lecture for you. For those of you who know something about it, I hope to at least impart some new knowledge that I have learned over the past couple of years.

I have been working in this field for about six years now and what I am going to tell you today is basically a bunch of knowledge that has been handed down to me from a lot of smart people and it is going to go through the Matula filter out to you. This Matula filter is very nonlinear, chaotic at times, and it is very noisy but, hopefully, I will be able to describe sonoluminescence to you.

[Transparency 2]

First of all, I will tell you what sonoluminescence is through a picture -- it is not this person, who was a graduate student of Larry Crum's at one time -- it is this little bubble -- right there. That little bubble is actually giving off its own light and that is called the "light fantastic," sonoluminescence.

The way it is achieved here is you have transducers. This particular cell has a transducer at the bottom and top of the cell, setting up a standing wave and you are filling the cell with water. You have to de-gas the cell, you have to remove a lot of the air out of the cell, about 10 or 25% of saturation, and you can inject the bubble in there by various mechanisms and you can turn up the pressure in the sound wave and the bubble will go to a pressure antinode. It is an antinode, so the bubble will expand and compress with the sound field.

As you turn up the sound field, the bubble gets a little jittery but soon after it locks into place and it actually gives off light. You can tell it is a bluish light.

The big questions after this phenomenon was discovered by Felipe Gaitan at the University of Mississippi, NCPA -- a couple of questions immediately came to mind. One is why is this bubble stable? You can have this bubble sitting in this flask, in the cell, for hours on end if it is properly tuned and it will not disappear. I will show you later that that should not be the case, this bubble should dissolve away.

The other question is what is the mechanism, why is it giving off light? Those are two questions I want to answer today that most scientists now agree on or at least feel somewhat comfortable with, agree as to what the mechanism is.

Before I leave this viewgraph, I need to tell you a story. This particular picture was published in Physics Today. Larry Crum was asked to write an article several years ago and this picture was published in Physics Today, only this bubble was missing in the published version. It turns out the graphics artist saw this smudge and erased it. (laughter)

[Transparency 2]

The graphic artist did not like sonoluminescence, but there are a lot of parts of society that have been enamored with sonoluminescence. A lot of newspapers have written articles on it, The New York Times, a lot of magazines have written articles on it, Popular Science, Physics Today, other magazines.

This one is the New Scientist. At the bottom here you can see "Bubbles Hotter Than the Sun." They have a couple of horns generating the sound waves here. This graphic artist did not like bubbles, either, so he just had a hand closing down and causing fire, or something.

The media, of course, has been very interested in sonoluminescence.

[Transparency 4]

Our government officials have also been interested in sonoluminescence. I actually had the ear of a Congressman for a whole half-hour, which is a long time for a Congressman, talking about sonoluminescence. He wanted to know how he could power his car with that glowing bubble, even after Willie Moss and I told him about a dozen times he would never get it done, it would never happen. Two weeks later we got another call asking how it would work and whether we could get his car powered.

The military is interested in it. This came from a Web site, I believe. Somebody got the ear of the Joint Chiefs of Staff and actually gave a talk on sonofusion, which down here says is related to sonoluminescence. Military officials have been interested in sonoluminescence, so the media and the government.

[Transparency 5]

Popular culture is interested in sonoluminescence. The First Annual Sonoluminal Conference: Music and Light. This is the light from sonoluminescence down here, so it is even in the popular culture.

[Transparency 6]

Of course, scientists are interested in sonoluminescence and what I am plotting here is the number of sonoluminescence papers published from about the mid-1960s on up to 1999 (I just got this off the Web). There has been a steady number of publications and I will talk about that in a second, but when Felipe Gaitan discovered single-bubble sonoluminescence, which is SBSL, the number of publications shot up dramatically.

[Transparency 7]

Most of the articles that you read about and most of the people who talk about sonoluminescence talk about single-bubble sonoluminescence and that is this area from 1990 on, but sonoluminescence was discovered, I believe, back in the early 1930s. There had been some work done in sonoluminescence and I want to move over to that area for a second.

If you take a high-intensity ultrasound source, you can generate sonoluminescence and that is that bluish light there. It is just thousands of bubbles and this is time-lapsed photography, just open the shutter for five minutes, and you get a bluish glow coming down from the transducer.

It is interesting that in the early 1930s, when this was discovered, it was difficult to see the light, so you can imagine what had to have happened. Some professor is telling his grad student, "Go into the darkest, dampest recesses of the lab, turn off the light for several hours, and see if you can see a glow." Of course, it was observed.

Here is another pretty picture of sonoluminescence coming from a high-intensity sound source. The transducer is located up here and you are getting a really pretty picture here.

I have in my notes here, to keep students awake, to ask you what kinds of parameters you could vary in this system to study. Does anybody want to volunteer? What kinds of parameters would I want to vary in this system to study sonoluminescence?

("Change the gas content.")

("Pressure amplitude.")

("Frequency.")

DR. MATULA: Very good. You guys are awake.

This is sonoluminescence, a bluish glow of light. It is related to sonochemistry. If you use high-intensity ultrasound sources, you can generate chemical reactions or you can facilitate chemical reactions. That is what a lot of people are interested in, especially in the chemistry community. They are interested in facilitating and enhancing chemical reactions in a particular process.

A lot of people wonder if you can use sonoluminescence as a probe. Instead of actually measuring the chemical compounds that are created, you can use just the light emission as a probe.

[Transparency 7]

I am going to go on a little excursion here out of sonoluminescence and talk about sonochemistry just for a couple of minutes. Using high-intensity ultrasound sources, you can change the chemical structure in your matrix. Up here we have CFCs, which was mentioned earlier in the week as kind of bad for the environment and they are being banned, but they can get in your groundwater supply.

What you can do is you can take some of this contaminated groundwater, put it in a flask, apply a high-intensity ultrasound source to it and you will get sonoluminescence but you also get a decrease in the amount of CFCs produced. This is the sonication time, concentration versus time. Over about a 10-minute period you can get rid of all these harmful CFCs.

Another application is for carbon tetrachloride, which is in the lower graph, concentration versus time, and this is in about 40 minutes. The concentration is in the red line. As you are sonicating it you are destroying the carbon tetrachloride.

Another example I have not shown here is parathion, which is an insecticide that has been banned. Parathion can be destroyed in about an hour; it has a half-life of 100 days, or something like that. Of course, when you destroy a compound you are forming other compounds. In the case of parathion you are forming other hazardous compounds but they are not as hazardous as the original parathion and that is good.

[Transparency 8]

Another interest is not just the chemistry that occurs but the formation of particles, in this case nanoparticles. In the very bottom left-hand part of the viewgraph here is nickel powder. It is very rough, it is under a microscope here (I forget what the magnification is). It is a very rough surface and if you sonicate it for about an hour, you can see that it gets very smooth.

In fact, here is a very interesting picture of one. It looks like two zinc particles have collided and fused together. It must be pretty high velocities of these particles to collide and fuse together like this.

You can make interesting particles. You can make amorphous iron using high-intensity ultrasound, which has kind of interesting applications.

MS. POLIACHIK: Is it velocity or heat?

DR. MATULA: I think people believe it is velocity. The person who gave me this viewgraph thinks it is the velocity of the two colliding. It is just particle-to-particle friction that is smoothing them out. That is a whole field all by itself, sonochemistry. There is a European Society of Sonochemistry. Everybody is interested in enhancing chemical reactions.

You can imagine a company like DuPont has a process and if you can apply ultrasound in the middle of the process to increase the chemical rate, your total time of processing decreases, your profits increase.

[Transparency 9]

Let's go back to sonoluminescence, though. I want to talk about the light emission itself. What I have is this handy-dandy little pocket spectrum that shows the electromagnetic spectrum from millimeters on up to gamma rays.

In the second column here is their origin and you can see, for visible light, which is what we are looking at, the blue light, you can see that the origin of this light comes from either outer electrons or molecular vibrations -- you can actually get molecular rotations.

The sources of these come from over here, either sparks, high-voltage sparks, combustion flames or thermal emissions. When sonoluminescence was first discovered, people thought it was an electrical phenomenon. You can be cracking the liquid open, generating a spark between the liquid fracture and the side wall containers.

Anthony Atchley, I think -- didn't you do your PH.D. study on the fact that cavitation comes from stabilized microbubbles and not from the fracturing of the liquid?

DR. ATCHLEY: That is too far back for me to remember. (laughter)

DR. MATULA: Anthony Atchley, I think, proved that it was not fracturing of the liquid. But even now there are people who believe that sonoluminescence comes from electrical discharges and it is reasonable; there are arc discharges and they give off visible light.

Most people believe it now is a thermal emission and I will show later that it seems to fit the data that it does appear to be thermal.

MS. PETCULESCU: But isn't it visible just because you can see lower and higher energies because of the water?

DR. MATULA: Yes. What happens is you see sonoluminescence and it is a bluish light, and that is in this range right here. We have actually gone down and looked in the infrared spectrum and you can actually see it decreasing down into the infrared, down about a micron -- that is right here, 1 micron.

You cannot go too far into the ultraviolet because water absorbs in the ultraviolet. I will show spectra later, but you can go to only about 200 nm in energy, that is about 6 eV or so. You cannot actually see at higher energies. If this emission were giving off x-rays, for instance, it would have been absorbed by the water long before it ever reached your eye, your detector.

[Transparency 10]

In multi-bubble sonoluminescence, when you have thousands of bubbles (and people have been studying this for years), you can actually calculate spectra. This is a particular carbon fluid, silicon oil, and they are looking at a particular emission, a Swan band emission. They can calculate these synthetic spectra, and I want to look at this <u>delta</u>-<u>nu</u>-equals-1 spectra right here. Basically, the calculation is by this formula that I wrote down.

There are a bunch of constants, <u>nu</u>, A, S, G, and F. Those are all constants that you look up in tables. They tell you what the energy of vibrations and rotations is, they tell you what the transition probabilities are. At a particular wavelength you can plug in those numbers and you can calculate the spectra at that particular wavelength and that particular temperature.

If you get an emission band, you can fit it to this synthetic emission spectrum and you can see what the temperature of sonoluminescence is. That was done in the late 1980s, I think, for multi-bubble sonoluminescence.

This was done by Flint and Suslick.

[Transparency 11]

The red line shows the measured spectrum (this is for the Swan bands, this is in silicon oil), and you can see the little bumps, shown here. For those of you who have seen spectra before, normally you have shorter wavelengths on the left and longer wavelengths on the right -- this is just backwards, so these are actually lower energy bands here.

If you take that synthetic spectra that was calculated on the last viewgraph, you can just match up with a single parameter fit temperature and find out what the temperature of sonoluminescence is, assuming that it is a thermal emission. This fit is for 4900 Kelvin.

For sonoluminescence in a multi-bubble field people have been measuring it now for 10 years and they always get some number near that, 4500, 3500; everything is less than 5000 Kelvin (the surface of the sun, of course, is also 5000 Kelvin).

One question I might want to ask you is why does the sun look yellow and sonoluminescence looks blue if they are both the same temperature? Think about it.

[Transparency 12]

Sonochemists have been studying sonoluminescence for a while now and they actually have a graph that tells you where sonoluminescence is. It is called <u>The Islands of Chemistry</u> and this is from Ken Suslick at the University of Illinois, Champagne-Urbana.

What he has here is the time of interaction occurring along this axis, the energy along this axis, and the pressures generated along this axis. You can see, for instance, that geochemistry occurs over very, very long time scales, relatively small energies but very high pressures.

Sonochemistry and sonoluminescence associated with that occurs way over in this region -- right here -- very short time scales, high pressures, and high energies. He likes to say (and I agree with him) that this is a physicist's dream, you can put all of chemistry onto one slide.

For the past several viewgraphs I have been talking about sonoluminescence from a cavitation field and sonochemistry associated with that. It is very easy to do chemistry from this field. These bubbles generate these chemical reactions, you can measure the chemical reactions from a lot of bubbles, but as a physicist this is very hard to study. You want to isolate a single bubble, you want to study a single interaction, and that we were able to do with single-bubble sonoluminescence.

[Transparency 13]

But now computers are getting powerful. This is a calculation of a multi-bubble cavitation field done by a Werner Lauterborn student in Germany. They can actually now calculate a million bubbles interacting and the qualitative nature of the cavitation field, so with computers now people are getting pretty good at being able to look at cavitation fields and not just a single bubble.

[Transparency 14]

But a single bubble has its advantages. It is a lot easier to study. Here is a bubble, again -- shown right there. I just glued a transducer on the bottom of it. This is mostly to show you two things: one, it is a single bubble (and this is what we want to study) and, two, you do not have to have a really nice cylinder to do your measurements in, you do not have to have a sphere or rectangular cell, you can use a beer glass, a wine glass, all sorts of different types of glasses. What you need is a three-dimensional standing wave that is stable.

One of the first things that was done when single-bubble sonoluminescence was discovered was to measure the radius of the bubble as a function of time -- it is motion.

[Transparency 15]

The blue dots on this graph show the radius as a function of time. What it shows is that during the negative part of the sound field the bubble grows, so you are just under tension, the bubble grows. When the pressure turns positive -- right here, somewhere -- the bubble cannot extend itself any further, it will collapse by inertia. By now it has grown by a factor of 10 or so, there is mostly vapor inside the bubble, it is just going to slam through the inertia of the fluid.

It slams down very hard, it actually gives off a flash of light there and I will show that a little bit later, and then it rebounds, like you have hit it with a sledge hammer and it is just oscillating at its resonance frequency.

At the very end of this sine wave, which happens just a little bit off this graph, the bubble's position ends up at exactly the same place where it started, so that when the cycle repeats itself, the bubble repeats itself in that exact same motion.

If you can get chemistry from thousands of bubbles, you should be able to get chemistry from this system. The problem is it is just a single bubble and you are not going to get very many molecules undergoing chemistry, but I have heard of a guy named Dudenko who actually is measuring chemistry from a single bubble.

I should mention something about how you make this measurement now. This bubble is growing and collapsing, so the way we make the measurement is we scatter light off the bubble. You scatter light off the bubble and you look at it with a photon detector, a photomultiplier tube. When the bubble gets big, more light gets scattered in the photodetector and, when it collapses, less light gets scattered into it. That is basically how you can get an increase and a decrease in the light scattered from the bubble.

It is actually a lot more complicated than I have described; something called Mie scattering theory is involved. But if you make assumptions, if the bubble size is big enough, you can actually just relate the scattered intensity to the bubble radius. I think the radius goes as a square root of the intensity.

MR. PORTER: Is it a laser beam?

DR. MATULA: Yes. What you want to do is when you shine the laser beam through the cell and it interacts with the bubble, you want to make sure the laser beam is much larger than the size of the bubble so that it is scattered from the whole of the laser beam.

The size of this bubble is 4.5 microns. It grows to about a factor of 10 times its original size. We are driving this system at a little over an atmosphere of pressure, so you have ambient plus the sound pressure of about an atmosphere. The frequency here is about 20 kHz, 20 to 50 kHz. That is where single-bubble sonoluminescence lives, in that frequency range.

In multi-bubble sonoluminescence you can take a high-powered sound source and you can operate it up to a megahertz if you want to, or higher.

You can scatter light off the bubble and you get its radius time curve, but now you will need to fit it to some equation that describes the bubble's motion. I am not going to drive it for you, I am just going to show it to you.

[Transparency 16]

It is called the Rayleigh-Plesset equation and there are a lot of different versions of this equation. A lot of them have terms in front of them that are different and that do not make sense but they are all basically equivalent. Basically, what it tells you is that there is an inertial part of this equation and it is being driven by the difference in pressure between the inside and the outside of the bubble, and if there is a pressure difference you will drive the bubble.

There is a very important part to that equation that I did not add in, and that is right here. The reason I did not add it in was just for simplicity, but I need to put it back in for me to tell you that there is no velocity of sound in the liquid in this equation and you need sound radiation to damp the bubble, and that comes from this term right here: It is the radius of the bubble divided by the velocity of sound in the water and the derivative of the pressure with respect to time.

That is a correction factor. That allows you to take into account sound radiation and removes energy from this oscillating bubble. There is going to be a paper coming out very shortly that actually makes another correction to this equation, which is 20, 30, or 40 years old.

When you think about the bubble oscillating, it is going to collapse and it is going to radiate pressure out into the water, but what is the most compressible part of the system? If you were going to make a compressibility correction to the system, would you make it the compressibility of the liquid or would you first put in the compressibility of the gas?

The gas. But people have neglected that part. If you can put in the compressibility of the gas, what happens is when the bubble starts collapsing, pressure waves get radiated inside. There will be another term added on to this equation including the compressibility of the gas. It will not affect the dynamics that you calculate very much but it is an important part of physics that has been neglected until now.

You can take this equation and put it in your MatLab program or your Mathematica program and for various values of the radius, surface tension, driving pressure, you calculate what we call an RT curve, radius-versus-time curve.

[Transparency 17]

Here are two such curves. The green is data and the blue is the fit, using that equation back there. You can see it works pretty well. The bubble here is actually not giving off light and it shows large after-bounces, but when the bubble expands dramatically, crashes, a lot of sound is radiated out into the fluid (that is why you need that term) and a lot of the after-bounces here disappear.

MS. PETCULESCU: So the key point here is the pressure?

DR. MATULA: It turns out that the bubble actually shrinks when you go up in pressure, so this bubble is at 10 microns and it is being driven at 1 atm. Then if you turn up the pressure, the bubble actually shrinks in size. I cannot really tell the bubble to be any particular size I want it to be. I will show later that gas diffusion governs how big the bubble is, but if you turn up the pressure the bubble will shrink a little bit and then the bubble will grow by a much larger factor.

This bubble is going from 10 to almost 30 microns. This bubble is growing from 5 to almost 40 microns. That is a problem in doing sonoluminescence research. Actually, when you are turning up the drive pressure, for a physicist, you want to change one parameter, say, the drive pressure, and look at the change in light emission, but what you are actually doing is you are changing the bubble size as well, and that was discovered several years after the fact.

It turns out, if you want to do good measurements, you have to actually measure the size of the bubble when you change a parameter.

Light scattering allows you to measure the size of the bubble over time, but there is another method that has come about recently.

[Transparency 18]

That method is just imaging the bubble. It is not too hard to do. You have a cell with your bubble glowing in the middle of it. You have a PZT driver that drives the sound field, it drives your standing wave, a certain frequency in amplifier to make sure you are tuned up, then you send a delay pulse into an LED and give it a short flash. On the other side you have a microscope and CCD camera.

What you are doing is you are looking at the shadow graph, basically, of the bubble. If the light flash is very short, you are basically freezing the bubble, you are strobing the bubble, or whatever phase of the sound field you want based on the delay generator.

Here are two pictures. This one shows a bubble at its maximum size (about 100 microns in diameter).

This one shows it at its equilibrium size before it starts to grow, and that is about 8 microns in diameter.

MR. APOSTOLOU: What is the bright spot in the middle on the picture on the left?

DR. MATULA: It is just the light going through the big bubble. You can think of it as direct rays going right through the bubble.

I am going to show that bubble growing and collapsing, using this imaging technique. [Video]

What this delay generator does is it sets the phase, so I capture the bubble at a single time in its oscillatory cycle, but if I put here another function generator and I synchronize their frequencies and then I just offset one function generator by a hertz or so different from the other one, then you have got this beating that is going on and the strobe is changing over many seconds and you can actually watch a kind of time-averaged motion of the bubble.

Here is the bubble growing and collapsing -- it grows, collapses, grows. You can see the rebounds that are occurring when it collapses. It is giving off light and if you turn off the lights, you can actually see the light, but now I think it is just being overpowered, probably by the room lights.

This particular set of data is two cycles in the motion and then repeated over and over, again. You can see after two cycles that the bubble has shifted a little bit. You can see that the bubble starts off at a different place after two cycles.

This bubble is very stable. You can carry it around with you, it is glowing. You can set it down, it will still glow, but it will drift a little bit.

PARTICIPANT: What causes that?

DR. MATULA: Thermal gradients in the liquid, things like that, just bulk flow. I will show you a little bit later that the bubble itself may be a little unstable, generating dipole motion of the liquid that will cause the bubble to move.

If you use a laser you can actually see the shock waves emitted from this bubble, you end up seeing a lot of refraction and diffraction and a lot more noise out in this area here. I will show a graph showing high-speed camera photography of the shock waves in just a second.

DR. COSTLEY: Did you change the timing?

DR. MATULA: No, what you change is, if you have two frequency generators, setting up a frequency between the two, you just change that. I just told the student to make sure the sonoluminescence cycle lasted, I think, four or five seconds, so he changed -- I do not know what the numbers are. You can make it last as long as you want, however close to a frequency matching you get.

MR. APOSTOLOU: So the inertial cycle is coming -- is it microseconds?

DR. MATULA: What I am doing now is I am averaging. This is not a single bubble going over one cycle in four seconds. I am averaging over hundreds of cycles and that is what imaging does. Your camera shutter is actually open for a long time and you are flashing and changing the flash mode, so it will not give you an instantaneous view of the bubble. Light scattering will.

What imaging does is it gives you a direct image of the bubble size, because you can stick a calibrated slide in there and you can know exactly what the size of the bubble is, whereas light scattering does not. It gives you an intensity and you have to work your way back to get a radius.

As soon as the bubble deviates from sphericity you run into problems. The bubble might jump out of place, it might eject a microbubble in a recoil. A lot of people believe (and the consensus is) that to get the light emission you have to have a geometrical collapse of the bubble, geometrical focusing of energy, so if the bubble is not spherical you do not get that geometrical focusing; it is squishy and you will not get a light emission, which is typically the case.

If I de-tune this system and I make the bubble jittery and bouncing around, the light emission is much weaker.

MR. DEMIRCI: Is there a difference between the top of the bubble and the bottom of the bubble, the pressure at the bottom of the bubble would be higher?

DR. MATULA: I will show that in the third hour, why we think buoyancy is an important parameter in this system. Under most calculations the bubble is only a couple of microns and the wavelength of this sound field is on the order of several centimeters, 5, 6, 7 cm, so the pressure gradient is not that big across the bubble wall, but I think it might be big enough to cause some sorts of instabilities and I will show that later.

You saw where the bubble collapsed. When the bubble collapses really strongly like that it is going to radiate a sound wave and that is why I had to add in that other term in the Rayleigh-Plesset equation.

[Transparency 19]

If you use a very high-speed camera, 20 million frames a second, you can see the shock wave emitted by this bubble. This is the bubble's collapse -- right here. Down here the bubble is still there and it is still collapsed. The shock wave is going off and you can see the shock wave expanding here as it goes out, and it looks like a spherical shock wave.

You can do another trick. You can scatter light off the bubble into a very, very fast photon detector called a streak camera. It allows you to look for a very short time interval at what is going on.

[Transparency 20]

When you do that, your laser beam will refract off the shock waves, and you can see that. That is what I show here. This was done by Pacha and Gompf in Germany. Here is the bubble -- this is time going along this axis. Here is the bubble collapsing, giving off light emission and then rebounding out here.

These two lines are the shock waves emitted by the bubble. I do not know if you can tell, but it is not a straight line, it is a bent curve, a shock wave that is being emitted from the system.

This is time, this axis down here, so you can imagine this bubble is collapsing right here, it gives off its light pulse and then expands out here, so that is expansion of the bubble. This is the expansion of the shock waves on either side of the bubble. That was just an image taken off the camera. This was done by scattering light off the bubble.

[Transparency 21]

Once you have these data you can actually plot the pressure versus distance. You can use the Tate equation of state for a liquid and you can plot the pressure as a function of distance. That is what is calculated here: pressure in kilobar or thousands of atmospheres. This is a 40,000 atm pressure as a function of distance from the bubble.

This decay is much faster than 1/R -- it would be interesting to see if it were a finite 1/R log R decay, I do not know if that is true or not, but it is faster than 1/R

-- and it loses about 50% of its energy in the first 25 microns. The calculated velocity of the shock wave is 4000 m/sec. It is very fast.

MS. HIGHTOWER: What is the key describing?

DR. MATULA: Different pressures. If you drive the bubble at a different pressure, you will get more intense shock waves or less intense shock waves. This one up here is being driven by a much harder sound field, it was collapsing much harder, generating higher shocks.

This one is not collapsing.

This triangle, circle, and other symbol represents a calculation of the pressure inside the bubble, a very elementary calculation inside the bubble, which seems to agree well with the measured pressure right outside the bubble.

Why don't we take a 15-minute break.

DR. MATULA: I was asked a very important question, and that is, how did Felipe Gaitan discover single-bubble sonoluminescence? I thought I would share with you the story. Felipe was using this particular cell at the time and he was studying multi-bubble sonoluminescence and the stability of single bubbles that were much larger but not sonoluminescent in a standing wave.

He would generate a cloud of bubbles in this area and turn up the pressure and get sonoluminescence. In another experiment he would just generate a big bubble in the center and study its dissolution and things like that.

One day, after spending all morning studying this one giant bubble and how fast it dissolved and all that, it was time for lunch, so he left and went to lunch. As a typical graduate student, he did not come back until evening. (laughter)

He walks into the room and said, "Oh, shoot, I left my equipment on." He looks at his apparatus and he sees the bubble is still there. "Oh, heck, by now it should have dissolved, it should have dissolved hours ago. Why is it still there?" He looks closer and asks, "Is it still glowing?" He runs over, turns off the lights, goes back to his apparatus, and says, "Oh, my God, it's glowing." He ran over to Larry and said, "Larry, it's glowing," and ran over to Ron Roy, "It's glowing," and nobody believed him for months on end.

So that is the story, but it is not the true story. (laughter)

That is the story I was told when I started in this field and that is the story a lot of people were told. When I asked Felipe directly, it was a coincidence. He was studying the phase of the light emission from a cloud of bubbles and every once in a while the phase locked at a particular phase in the sound field and it took him a while to figure it out, but it was a single bubble.

The part about Larry Crum not believing him is true. It took him a lot longer to convince his adviser that it was as single bubble. (laughter)

What he believed happened is that the water, the fluid he was using, which was a combination of water and glycerol solution, was too gassy to sustain single-bubble sonoluminescence. You have to de-gas the liquid, but if you are cranking up the pressure and you are studying a bunch of bubbles, over time it is going to de-gas itself, so he thinks over time it de-gassed and just by happenstance was able to get to the right pressure regime to generate single-bubble sonoluminescence.

MR. PORTER: I have a question. How do you get just the one single bubble?

DR. MATULA: What you do as an experimentalist is you can inject air into the system somehow. A lot of people use the Hiller method, which involves taking a little nichrome [phonetic] wire, putting it inside the fluid with the water, and putting heat through it, boiling the water around the wire and generating vapor bubbles, and then those vapor bubbles will be attracted to the pressure antinode and on the way there they will fill in with gas, but that is lots of vapor bubbles and all sorts of different sizes.

Nature decides to change the size through gas diffusion to whatever size it likes best.

PARTICIPANT: So these are multiple bubbles that coalesce?

DR. MATULA: They will coalesce into a single bubble, yes. If you have stabilized bubbles, like contrast agents, they will agglomerate but they will not coalesce. I have done this with contrast agents and you get a whole bucket of them all around but they will not coalesce into a single bubble.

[Transparency 22]

Felipe discovered that it is the phase of the light emission that was very important, and here is the phase. You have the sine wave driving your system and at a particular phase in the sound field you see a single light emission going at a particular phase. If you have multi-bubble sonoluminescence, it is scattered all over the place, but a single bubble locks into a particular phase.

Seth Putterman at UCLA and his graduate students did a really nice set of experiments to show what the period between successive flashes was. You are taking a flash, you are looking at the period to the next flash -- it is about 50 μsec or so -- and then you are looking at the jitter in that period. They are finding that the full width at half-max is on the order of 100 psec, so it is a very synchronous event occurring at the same time cycle to cycle to cycle.

Anthony Atchley and a collaborator, Glen Holt, found that if you de-tune the apparatus a little bit you can actually get period-doubling effects and continue on into chaotic events, where the synchronicity breaks down and you get just chaotic behavior but, if it is extremely tuned, you get a very synchronous event, and that is the mark if single-bubble sonoluminescence.

[Transparency 23]

One of the most important measurements that was made is measuring the pulse width of the light flash. It occurs once every acoustic cycle but it is a very, very short flash and it is hard to measure the width of that flash.

Now there are a couple of methods to do it. This was the original method. Basically, you have your flask and you have a little bubble in there glowing and it gives off a flash of light, so a photon may leave this side and reach this photodetector and another photon, some time during that flash, may leave this side and enter this photodetector and you use filters to limit what wavelengths you want to look at and to make sure that you are looking at only single photons.

The photons go through a timing device, a constant fraction discriminator (it is just a timing device) and then the two photons go into this TAC, a time-to-amplitude converter, which basically gives you a voltage level proportional to the time between when it saw the two photons.

If you do this with a femtosecond laser, they get this curve right here, which shows that they can resolve the pulse width of a femtosecond laser within 49 psec, so that is basically the impulse response of their system.

When they look at single-bubble sonoluminescence, they get a curve that looks like this and they actually measure counts, because you build up a histogram here, and they measure a pulse width on the order of 100 psec. So the pulse width of sonoluminescence is on an order of 100 psec. It has been measured to down below 50 psec and up to around 400 psec, or something like that.

Another important feature of the sonoluminescence light flash is the fact that it is asymmetrical; it does not go up and come down symmetrically, it has an asymmetry to it. This particular method of measuring sonoluminescence flash widths cannot resolve which side the asymmetry is on, so when this paper was published in Physical Review Letters they had some physical argument as to why this asymmetry should actually be on the left-hand side here but, over time, people realized that it should actually be on the right-hand slide, so I flipped their diagram over for them.

PARTICIPANT: Then this is some type of artifact?

DR. MATULA: No and, in fact, there are lots of artifacts that they have to get rid of. When photons leave the bubble they will reflect off the glass, you will get reflections off the glass so, actually, you end up with bumps on either side of these that you have to get rid of. You have to worry about that kind of stuff.

MR. APOSTOLOU: I do not understand. Why can't you use light detectors inside the medium? Obviously, I do not know if there are any, but --

DR. MATULA: A light detector inside the water?

MR. APOSTOLOU: Yes.

DR. MATULA: To do what?

MR. APOSTOLOU: To measure the light, because you mentioned the problem with the glass walls of the container, and so forth.

DR. MATULA: I suppose you could take like a fiberoptic, stick it next to the bubble, have the light go in, and then do a fast photomultiplier tube.

MR. PORTER: If you had a standing wave, would that do any disrupting?

DR. MATULA: If you had a really thin fiber you could probably get it pretty close to the bubble. For this particular method you need two photons, you need two detectors, you would have to put two fiberoptics in there. Actually, aligning a fiberoptic over the bubble is not easy to do. Some people do spectra from the light, because you want to measure the energy from the light, so they stick a fiberoptic over there and if you are off by just a fraction of a millimeter, you are missing everything, because a solid angle is very important.

DR. CRUM: At that time you get a lot of dispersion, so you will spread the pulse way out.

DR. MATULA: So all fibers are that dispersive?

DR. ATCHLEY: At this scale. You also have problems with phosphorescence and things.

DR. MATULA: So this experiment, of course, is really nice. In fact, what you can do with these filters here is you can put a red filter, a blue filter, you can pick out the energy levels that you want to look at from the light flash.

If sonoluminescence is blackbody radiation, then, when the bubble collapses, it will get hot in a certain order: infrared, it will start emitting red light, then green, yellow, blue, and then, as it cools down, the last part of the light that will be emitted is red. If you put red filters up here and then put blue filters, you can compare the length of the emission. If it is blackbody radiation, the red pulse width should be longer than the blue pulse width.

It is exactly the same if you do the measurement. There is no difference between red and blue at these temperatures, at normal ambient temperature, so that immediately ruled out sonoluminescence being due to blackbody radiation.

That immediately ruled out a lot of other things, too. If you assumed it was a blackbody radiator, it was a very hot blackbody radiator.

[Transparency 24]

Here are measurements of the pulse width as a function of how hard you drive the bubble. Remember, if you drive the bubble harder, you get more light. Also, what happens is the pulse width increases as you increase the driving pressure. These are three different experiments for three different concentrations of gas inside the liquid.

As you de-gas the liquid, you are going from top to bottom. As you remove gas from the liquid, a particular driving pressure will give you shorter and shorter pulse widths. As the UCLA group asked, why are you stopping here, why not de-gas the liquid more and more?

They did and they ended up with pulse widths on the order of 40 psec, so you can actually get lower than 50 psec for a pulse width. That is a very, very short pulse width.

[Transparency 25]

Another important parameter to measure is the spectrum. Any time you see light being given off from something you want to measure its spectrum. These are some of the first measurements done by the UCLA group. This shows a spectrum of different gases dissolved in a liquid.

This purple one is argon. This is a log-log scale, the radiance watts per nanometer and the wavelength from the UV out toward the near infrared. In argon, for instance, it continues to grow into the UV. Helium does the same thing. In fact, most gases will continue to grow as you go out into the UV until the cutoff of water, or the cutoff of your quartz flask, which starts to absorb too much.

It turns out that the cutoff of water is actually 180 nm, not 200 nm, so I am not exactly sure why you cannot go a couple of nanometers farther, but there might be some errors that just grow exponentially there.

There are some gases that show a turnover, a maximum, in the light intensity at a particular wavelength. Those gases happen to have xenon in them, or mixtures of xenon.

MS. PETCULESCU: For those pure gases, how can you do that? How can you make sure you have only argon?

DR. MATULA: The way they do it is they take their water, they filter it, de-ionize it (I am pretty sure this is what they do), and then they attach a vacuum to the head space above their container of water. What that means is they remove all the air above the water, so then the gas inside the water will move up so that there is an equilibrium in concentration above and below the water surface.

They will keep on pumping out the gas until they have removed as much air as humanly possible. There is always going to be some air, you cannot an infinite amount of air out of the system, but they will de-gas the system as completely as they can.

Then they will pump in whatever gas they want to pump in, argon, say, and they will pump in argon, they will stir it in, they will saturate the thing with argon, then they will de-gas it again. Then they will pump in more argon and de-gas it again. They will go over and over, again, until, experimentally, they see no differences when they do an experiment.

They might refill this water with argon four or five times before they actually do an experiment. Nobody can guarantee that there is not a single air molecule in there but it is almost all argon.

MR. APOSTOLOU: Then they just inject the bubble with argon?

DR. MATULA: They probably use a nichrome wire approach, where they have this little nichrome wire in the cell and they boil it. The problem with injecting bubbles is you might contaminate the system with air, or something.

You can get whatever gas you want into your system just by backflowing it in under pressure.

MR. PORTER: What is the best combination in terms of gas? 2% of xenon in nitrogen or --?

DR. MATULA: The amount of xenon you put in, or the amount of argon you put in, is interesting. You end up having to put in a certain amount to make the bubble stable. You cannot just inject any bubble at any gas concentration and make it stable; only a certain range of gas concentrations make it stable. These particular concentrations give a stable bubble.

These spectra do not show band emissions as I showed for multi-bubble sonoluminescence, so you cannot fit a temperature to it very easily. You can take this as a tail of a blackbody radiator, even though I have already shown you that it cannot be blackbody, and take a fit to that and see what the temperature is, and people did that.

[Transparency 26]

Here is the spectral radiance on the side here versus wavelength, from 200 to 700 nm. They fit the data, shown as blue squares, with this formula and they fit it with a single temperature, and they had to fit the wavelength, too, here.

They got temperatures, if you believe this, from 1000 to 500,000 Kelvin. It is very hot. When these data were taken and this information was released, immediately nuclear physicists went into their codes and showed that the bubble was not only 500,000 Kelvin, it was on the order of millions Kelvin. It was a very hot bubble.

PARTICIPANT: How long does it stay hot?

DR. MATULA: It stays hot for picoseconds.

At the time, five years ago or so, this is what people thought was sonoluminescence: an extremely hot phenomenon and maybe even fusion was occurring.

[Transparency 27]

During this time there were these questions that were asked. Why is it stable? I told you at the very beginning that it should not be stable, it should dissolve. What was the mechanism for light emission? I have already told you it is not blackbody radiation but, still, what is the mechanism? Multi-bubble and single-bubble sonoluminescence have different spectra. Single bubble does not show band emissions, but multi-bubble sonoluminescence shows band emisssions, so does that mean they have different mechanisms?

Finally, can Iraq foil the nuclear test ban treaty? Why do I say that? Well, these are very high temperatures, right? This could be millions of degrees in temperature.

[Transparency 28]

The London Independent wrote an article on how Iraq could get around the International Atomic Agency and test nuclear fusion -- I believe it says "sonoluminescence" on there -- that is one way they could get around it. This was a time period when people really thought sonoluminescence was very hot, so when you are offered to go look for neutrons underneath a mountain in Provo, Utah, you take the opportunity and go.

[Transparency 29]

When I was asked to go look for neutrons in Provo, Utah, I took my camera and my sonoluminescence rig and I went under a big mountain. You go under a big mountain to look for neutrons to get rid of cosmic rays that increase the background radiation.

You add salt bags around your apparatus to get rid of cosmic rays. You add veto counters. These are basically fluorescence devices that are attached on the outside that are used as timing devices. If it receives a signal and your experiment has a signal and they are correlated in time, then you know it was a cosmic ray that caused it.

I spent a week trying to find neutrons from single-bubble sonoluminescence, because it is a very hot phenomenon and you should see neutrons. It turns out there are no neutrons that I could see from the system, but I had a good week.

[Transparency 30]

Just because I could not find neutrons does not mean Hollywood cannot find neutrons. In fact, they had a movie, called "Chain Reaction," and I hope none of you saw it. The nice thing about it is three square blocks of Chicago got blown up in their sonoluminescence apparatus and though I would not mind being a part of that, I do not think it could happen with me. (laughter)

That is me playing Keanu Reeves. Okay, okay.

So we are answering one of the questions: Can Iraq foil the test ban treaty? Hollywood says yes, they can.

Another question we wanted to ask was the difference between multi-bubble and single-bubble sonoluminescence. In multi-bubble sonoluminescence you see band emissions, you can calculate its temperature, it is 5000 Kelvin.

[skip to Transparency 40]

In single-bubble sonoluminescence there were some calculations showing it to be 100,000 Kelvin or so. I will show later that most people now believe it is on the order of 10,000 to 20,000 Kelvin, so it is slightly hotter than multi-bubble sonoluminescence.

We did an experiment in sodium chloride solution, just water with table salt. We measured the spectra in single-bubble sonoluminescence and multi-bubble sonoluminescence. In single-bubble sonoluminescence you get this blue line and it is pretty much a continuum from low energies to short wavelengths for high energies.

If you do the same experiment in multi-bubble sonoluminescence, you see this telltale sodium emission line at 589 nm.

How do you get sodium inside a bubble? Sodium is nonvolatile, so how do you get it inside the bubble? In single-bubble sonoluminescence I would expect to see a continuum, because I cannot imagine how you get sodium inside the bubble.

In multi-bubble sonoluminescence people have had several arguments for it, but one of the accepted ideas is that you have lots of bubbles interacting with each other, generating jets, as Larry Crum showed this morning, you have jetting going on, so you can actually get fluid inside the bubble and, therefore, you can entrain sodium inside the bubble and then you can get sodium emissions.

Another possible effect is that you can actually get a shell of liquid surrounding your bubble hot enough to cause sodium emission, so why don't I see it in single-bubble sonoluminescence? It might be that I just do not have enough sodium in the solution to be able to see it. You cannot just continually add a lot of sodium because the bubble becomes unstable.

The third mechanism, there is another mechanism by which you can get sodium emission; that is, you can generate radicals inside the bubble that leave the bubble and interact with sodium atoms and you can get sodium emission through radical chemistry, which I know nothing about.

So the difference between single-bubble sonoluminescence and multi-bubble sonoluminescence is quite dramatic when you look at the spectra.

[skip to Transparency 41]

But it turns out that you can get interesting effects with single-bubble sonoluminescence if you go to another liquid. Until very recently everybody doing single-bubble sonoluminescence has done it in water. Ken Suslick's group has been able to use another liquid and they have been able to get single-bubble sonoluminescence, stable single-bubble sonoluminescence, and measure the spectrum from the system.

This blue line is the spectrum from a stationary bubble and these wiggles are just noise. If they then change the pressure or de-tune the system so that the bubble is no longer stationary but is moving around, they actually get a bump in their spectrum and it corresponds exactly at the right wavelength for CN emission.

What scientists think what is occurring here is that when you have a stationary bubble you have this bubble growing and collapsing and you get geometrical focusing and the bubble gets hot, but as soon as it becomes jittery, you cannot focus as much energy in the system, the bubble will not get as hot and then you can see these bound-bound transitions generating band emissions. These are vibrational CN transitions.

DR. KEOLIAN: What was the liquid?

DR. MATULA: I was waiting for somebody to ask that question. I cannot tell you. This has not been published yet.

DR. CRUM: It has been accepted at Nature.

DR. MATULA: Are you sure?

DR. CRUM: Well, it got one good review and I know another is sent. (laughter)

DR. MATULA: You tell them, then. I was told I cannot tell them. What I am told is it is a liquid that physicists do not want to use, that it is dangerous, but it is really not that dangerous as long as you are careful with it.

PARTICIPANT: What is CN?

DR. MATULA: Carbon nitrogen. I do not know what it is called, the name of it.

So multi-bubble sonoluminescence appears to be a weaker, cooler form of single-bubble sonoluminescence and that is what a lot of scientists now think is occurring, though not everyone. I was just at a conference in Europe where at least one scientist still thinks it is an electrical discharge phenomenon occurring, but most of us believe it is a thermal effect.

[Transparency 31]

Two other questions: Why is the bubble stable and what is the mechanism of sonoluminescence? That is what I want to answer. I want to answer, first, why is the bubble stable over hours on end?

To answer that question we need to look at diffusion across the bubble wall. If the gas diffuses into the bubble faster than it diffuses out, the bubble will grow. If the gas diffuses out faster than it can diffuse in, it will shrink over time. That is given by this diffusion equation.

You also have the bubble's motion, which is governed by the Rayleigh-Plesset equation, which is shown here. These two equations are actually coupled. If you change the gas concentration inside the bubble, then that has to be related by the pressure. If you change the pressure and drive the bubble harder, you are going to change the gas diffusion.

Although nowadays you can actually solve this with powerful computers, it is a lot faster just to make some approximations and get rid of, for instance, the convective term there.

[Transparency 32]

So what people do is they make some approximations. They assume that the radius time of the bubble is periodic, which it is, every acoustic cycle is periodic. They also assume that you do not care what is happening over a single acoustic cycle; what you want to know is whether this bubble is stable over long time periods, whether it is going to grow or shrink, so you ignore what occurs in a single cycle.

You go through a lot of mathematics and you end up with an equation that relates how many moles of gas are entering or leaving your bubble and it depends on the radius time curve of your bubble, and it depends on how gassy your liquid is. This ratio tells you how much gas you have put into the liquid divided by how much it can sustain saturation concentration.

What you do is you take that Rayleigh-Plesset equation and you solve it in MatLab or Mathematica, you plug it into here, and you can get whether dN/dt is negative or positive, telling you whether the bubble is growing or shrinking.

MS. SWEARINGEN: You said that it is definitely periodic but in the video you showed us it looks as if it collapses much faster than it grows. Is it just because it is collapsing at the same rate over time and growing at the same rate of time that it is periodic even though it is not evenly growing and shrinking?

DR. MATULA: Right.

[return to Transparency 15]

This bubble is going over one acoustic cycle here, just a little bit less than one acoustic cycle, and it is doing this motion, growing, collapsing, rebounding. Then at the very end it repeats itself, so that is the periodicity of it.

There are actually a lot of different time scales involved here. One is the period of the sound field. One is the growth cycle that occurs over about half a period or so. The collapse phase occurs over a couple of microseconds. The light emission phase occurs over picoseconds and the rebounds occur over several microseconds, so there are a lot of different time scales involved in this phenomenon.

MR. APOSTOLOU: Why is there a time lag between the pressure, the maximum and the reduced?

DR. MATULA: Inertia.

[Transparency 32]

You can take this equation, and I think I found a mistake in it already. What happens if $R=R_0$? You get 1, right? What happens if you get $C_{infinity}$/CR are equal? Take a 4-micron stationary bubble, saturate your liquid so the saturation concentration, $C_{infinity}$, is equal to thesaturation concentration and this equation predicts that your bubble is stable.

What is going to actually happen to your bubble? It will dissolve. Actually, you have to include surface tension here, which I did not include.

Take that equation, plug it into MatLab along with your Rayleigh-Plesset equation, and you can generate curves like this.

[Transparency 33]

For various values of your equilibrium radius and various values of your pressure you can determine whether dN/dt, the number of molecules entering or leaving the bubble, is positive, negative, or zero. This red line corresponds to the equilibrium condition, where the bubble is stationary. This region inside is the growth region, where more molecules are going into the bubble over time than leaving the bubble, so the bubble will grow, and the region outside is the dissolution region, so there are more gases leaving the bubble than coming into the bubble and the bubble will dissolve over time.

Let's take a look at this, because this is very important. A sonoluminescence bubble should not be stable and it is based on this diagram, so I am going to go over it with you, so pay attention.

If you are driving this bubble at 1.35 atm and it is about 5 microns, you are right on that line, so you are stable, the bubble is not growing or dissolving. Suppose the bubble undergoes a hiccup and grows a little bit to 7 microns. What is going to happen to it? It is going to dissolve. It is going to come straight down this line because you are driving it at a constant pressure.

What if it undergoes a hiccup and goes down here? It is going to grow. So not only is this an equilibrium but it is a stable equilibrium.

You can have a bubble down here, too. It can be right on this line and it can be in equilibrium and glowing, but if it goes off this line in one direction or the other, it will continue to move off in that direction.

This line, of course, not only depends on the RT curve, the Rayleigh-Plesset equation, but it depends on that ratio of how much gas is dissolved in your liquid, $C_{infinity}$/CR. For this line that ratio is .005. If I chose .002 it would be down here more. If I chose a bigger number, it would come out here farther.

MR. GLADDEN: I am sorry, would you tell me again what that pressure is?

DR. MATULA: The drive pressure, how hard you are cranking the system. Pressure amplitude. It is a sine wave.

MR. TUTTLE: Does the bubble oscillate between those two states of it?

DR. MATULA: I see it all the time. Bubbles oscillate between states all the time. I am not sure if this is the reason for it or if it is the fact that in a real system, when you set up a standing wave in a system, you are actually also generating higher harmonics. If your three-dimensional standing wave is not very stable, the bubble can actually jump between different positions, actual physical positions. That might be related to this, I do not know.

If you went from this state up here and hiccuped down to this state, it is going to continue to dissolve unless it is stabilized somehow by moats or crevices, or something like that.

[Transparency 34]

This is the calculation. Now we can apply it to a sonoluminescence bubble. I like to use R_0 as a function of Pa, the equilibrium size of the bubble as a function of the driving pressure, but the group that did the experiment actually likes to go Pa versus R_0 differently. Maybe what I could do --just reverse the axis, pressure amplitude as a function of radius. These are measured radiuses of bubbles. This is the equilibrium line.

This is the sonoluminescence bubble up here. These bubbles up here are being driven at a little over 1.25 bars or atmospheres. Their sizes are 2, 3, 5, 7, microns in radius.

These other bubbles are bubbles that are unstable. These are all unstable bubbles over here. If you generate a 5-micron bubble at 1.15 or 1.2 atm, it is in a region of growth and it will grow out to this instability line and it will undergo instabilities, and you can see that in a video camera and you say, okay, that bubble is unstable right there, so that is how this line was made.

This line right here was calculated through that diffusion equation I showed you in that other calculation to match the sonoluminescence bubbles that they had in their system.

These are the actual values that they found -- right here. If sonoluminescence bubbles obeyed gas diffusion, then this line -- they should have been along either this line or this line should have been along the sonoluminescence bubbles, one or the other, so there is obviously something wrong here. We are not predicting the correct gas diffusion that is going in single-bubble sonoluminescence.

It turns out the answer to that was given by a physicist and all the chemists are mad at him for this reason: When the bubble collapses, it gets hot. What happens when you have a hot bubble? You have chemistry. You have air inside your bubble, you have nitrogen, oxygen, and some other contaminants. The nitrogen and oxygen will undergo dissociation when it is hot. They will dissociate, they will be excited, they will form other products when it cools down, like nitrites and nitrates, NO_X compounds. Those are readily dissolvable in liquid, they will migrate out of the bubble and dissolve out into the fluid.

What will not generate chemistry is argon, a very, very small percentage of air, but it is argon, so that will stay in the bubble. It turns out, if you take this calculation for air and you say "I don't have air in my system, I have, actually, argon, which has a concentration a hundred times smaller than air," then this line will shift right along there.

[Transparency 35]

Bob Apfel's student, Jeff Ketterling, who has now gone off to bigger and better things, actually did those measurements. This is back along the way I like it, which is R_0 is a function of Pa. Here is the gas diffusion equilibrium curve for pure argon saturated to 0.26% of saturation.

He generates bubbles and they kind of follow along that curve, so when you have pure argon the bubbles seem to follow along the curve. You get stable bubbles and even some unstable bubbles of the right size being driven at the right pressure amplitude.

Then you take pure nitrogen. This is a curve corresponding to pure nitrogen saturated to 10%. For a particular bubble size at a particular driving pressure you can get bubbles. These particular bubbles are not glowing but they are stable and they should fit on this diagram.

They fit pretty close; they kind of go up where they are supposed to go up, but not like that other graph I showed you.

If you take 1% of argon dissolved in 99% of nitrogen, what is that? That is air, basically, mostly nitrogen, some oxygen, and 1% argon. You dissolve it in your system. You have two curves, depending on whether you think your system is full of argon or full of nitrogen.

This curve corresponds to nitrogen.

This curve corresponds to argon.

If you are driving the bubble and it is not giving off light, you get bubbles that are stable along this curve right here, but if you crank up the pressure and the bubbles start to emit light, what happens is they all shift over to this side of the curve, which is an inference that what has happened is you have had dissociation of the nitrogen leaving the bubble, leaving you with argon, so the bubble is now an argon bubble.

[Transparency 36]

Another experiment was done where we looked at the evolution of the light emission by the bubble below and above the sonoluminescence threshold, below, where it is not glowing, and then above, where it is giving off light. That is the blue marks. The red line is a timing device, basically.

On the left side of this timing vertical line the bubble was stable, not giving off light. We then cranked up the pressure amplitude really fast to a state where it would be giving off stable sonoluminescence. Of course, it is a finite system; it will take some time, depending on the Q of the system, for it to respond and for the bubble to give off light.

But if you notice here, it starts to give off light, but very slowly. It actually takes about 10 seconds for the light to become stable and bright. If you take that stable bright bubble giving off light and you then bring it down below the sonoluminescence threshold, when it is not giving off light, you get this intensity, which is zero, again, but then if you immediately bring it back up to give off light again, it immediately comes up and, instead of having this delay, it immediately gives off very bright intense light.

Basically what we argued was that there is a 10-second time frame that chemistry and gas diffusion is occurring over for the bubble to become an argon bubble. If you bring it back down to not giving off light and you leave it there long enough, the gas will diffuse back in and it will become an air bubble, again. If you do not, if you bring it back up really quickly, it will remain an argon bubble and will emit light right away. It is about a 10-second transition phase for the chemistry and the gas diffusion to occur over.

Based on those two sets of data, people are fairly comfortable now, at least with noble gases, nitrogen, and oxygen, to say the following: When you start off with a bubble and it is undergoing sonoluminescence you are going to get dissociation of the diatomic molecules, whatever is in there.

If they are miscible in water, they will dissolve out into the fluid and what is left over will be whatever nonreacting species you have, which, in this case, is a noble gas, so what you think is an air bubble when you added air into your system turned out to be an argon bubble.

[Transparency 37]

Sonoluminescence exists in a particular parameter space. I said before that at a particular frequency range we get it between 20 kHz and 50 kHz. Why don't we go below 20 kHz? I have gone down to 7 kHz -- not me, personally, I had a student do it -- but why would you not want to go down to 7 kHz? It hurts. It hurts big time, so 20 kHz, the limit of human hearing, people do not want to go below that.

People have gone up to about 70 kHz, but after you get up to higher frequencies the Q of the system becomes such an important factor that it is hard to generate a stable bubble.

In pressure amplitude there is also this region of stability. At very low pressure amplitudes you just have an oscillating bubble. If you drive it harder, it actually goes into this dancing motion, nonstable behavior, and if you keep on cranking it up, you generate a single-bubble sonoluminescence.

You have a lower threshold and you have an upper threshold, this extinction threshold. If you drive it too hard, you drive it above this threshold, the bubble disappears.

[Transparency 38]

It turns out that if you drive it near that threshold you should start seeing stability problems. The stability problems are observed by watching the bubble jitter around the flask but are calculated by assuming you have a spherical bubble and you are going to superimpose a nonspherical oscillation on it, spherical harmonics.

If you remember spherical harmonics, they have a coefficient out in front of them. If that coefficient is zero, you have a spherical bubble. If that coefficient is greater than zero, you have some other mode that is not a pure sphere. So you can add spherical harmonics to the system to induce quadripole oscillations or higher mode oscillations, if you want, and you can look at the evolution of that factor in front of the spherical harmonic and see whether it grows or shrinks in time, and that kind of tells you, calculationally, whether your bubble is going to stay stable or not.

This is a calculation, three different conditions, for a bubble where you have added on a spherical harmonic, where you have added on an instability. This instability -- right here -- is calculated to occur every acoustic cycle and it turns out it occurs right when the bubble is collapsing and people call that the Rayleigh-Taylor instability.

You also have an instability that occurs every cycle but kind of grows and shrinks over time and kind of goes back and forth and that instability these researchers, Hilgenfeldt and Lohse, have termed "afterbounce instability"; it occurs right when the bubble collapses and then a few afterbounces afterward.

Then there is the parametric instability that grows cycle to cycle to cycle. All of these instabilities are calculated to occur in single-bubble sonoluminescence and all those instabilities can extinguish the bubble, can destroy the bubble.

[Transparency 39]

If you take those instabilities that are calculated, you take the gas diffusion equation and you put them all on one plot, this is the region where sonoluminescence exists. You have R_0, the size of the bubble you are starting with, that nature determines (I cannot determine that for it, nature determines the size of the bubble), the driving pressure amplitude -- that I can fix -- and what you have here is the following.

Let's look at this red line. This is an Mach=1 line. At some point, the bubble, as you increase the pressure amplitude, starts to emit light. We do not know what that lower threshold actually is but we assume that the bubble must be going at least M1 or so, the speed of sound in gas, to be able to emit light. We call that the lower boundary. The bubble has to be going faster than that.

On the other end, the bubble cannot generate instabilities, so the upper boundaries are these instabilities. Here is one right here, going across the top, and here is one, right down, going down. Those are those instabilities I showed on the last viewgraph.

Sonoluminescence occurs in the region between the instabilities and this lower M=1 threshold. It does not occur anywhere in here. It actually occurs along the line and that is how much you have de-gassed the water.

This curve tells you where sonoluminescence will exist in your system. The idea, of course, is to expand this region as much as possible. We cannot expand it going down this way, because the bubble has to be collapsing at a certain velocity before it gives off light, to begin with.

The idea is to expand this region up here, or this region on the side here. If you expand those regions, you are going into larger bubble sizes, if you can go up here, which gives you more time to make measurements when the bubble is collapsing. If you expand this region out here, you can actually drive the bubble harder.

The point I want to make for the second hour is that there is a parameter space that single-bubble sonoluminescence lives in. You have to take into account chemistry that is going on, because chemical reactions are occurring, but if you take those into account, you can define the parameter space that sonoluminescence is in.

In the next hour I will talk about what makes sonoluminescence, what that is.

DR. MATULA: I described why a bubble is stable and that involves chemistry. Now I am going to describe the mechanism of sonoluminescence. It is still somewhat controversial -- remember, you are going through the Matula filter -- but it is the most accepted theory of the mechanism of sonoluminescence.

[Transparency 42]

The assumptions here are that when you compress the bubble, it gets hot, and when you have a hot bubble it emits light. Those are assumptions, because other people will say, "My assumption is that when you compress a bubble, you have charges on the outside and those charges generate an electric field and the light emission comes from the discharge." So this is a different assumption. This is a thermal assumption.

Basically what happens is that when you compress the bubble you get exciting molecules, they dissociate. You get sonochemistry from this (those first couple of viewgraphs that I showed you). You get partial ionization that is calculated to be less than 10% -- it depends on the noble gas that you use. At the bottom of the collapse or very near the bottom of the collapse you start emitting photons and then the bubble rebounds.

[Transparency 43]

There are a couple of different ways to look at this. One way is to take your bubble motion, send it to Lawrence Livermore National Laboratory, have them take their bomb codes and see what happens inside the bubble. When you do that and you make a couple of assumptions, you come out with a formula that describes the radiating power from a sonoluminescence bubble.

This is the mechanism. The bubble is a combination of a thick and thin radiator. What do I mean by that? There is an optically thick, or opaque, core that you cannot see into. It is optically opaque. That means that photons generated in the core are absorbed immediately, so they do not come out. It is kind of like our sun; photons generated in the core of the sun we do not observe.

Then there is an optically thin region surrounding the core -- this is the optically thin region, the actual bubble size is much larger than that. It is a shell, it is thin, it is not opaque, you can actually see into that region and that is where you see the light emission. It is like the shell of the sun, why we think the sun is 5000 Kelvin when it is actually generating thermonuclear fusion.

The most important factor that people discovered in adding in new physics to try to figure out why single-bubble sonoluminescence was not a blackbody radiator was this term right here, Kappa, the opacity of the radiating matter. The opacity basically tells you how far you can look into this bubble.

I drew out this morning what might happen with a xenon bubble and an argon bubble as you go into the bubble along this axis (or that could also be time), and this is the temperature. If temperature is increasing, the opacity for xenon is different from the opacity for argon. The opacity, of course, is photon absorption.

The opacity for xenon will turn on at a lower temperature, which means you will see it for a longer period of time and over a larger volume than you will for an argon bubble, where the opacity turns on at a higher temperature, which would be at a smaller volume.

For this mechanism the radiating volume is the important aspect that gives off light emission. Xenon looks brighter than argon, not because it is hotter than argon (or it may be), but mostly because it has a larger radiating volume. It turns on at a lower temperature and it emits light for a longer period of time and it looks brighter, more photons.

[Transparency 44]

You can calculate a spectrum with this formula. This was done by Willie Moss at Lawrence Livermore National Lab. You get the following. This is the spectral energy density and notice the units here, femtojoules. The amount of energy given off by this bubble is not enough to power your car. I will bet more energy is given off by a mosquito slamming into a wall than is given off by sonoluminescence bubbles. This is as a function of wavelength.

The data points for argon are shown as blue circles here, the circles with the blue line, and the data points for xenon are shown by the circles with the red line going through. The calculation depends on the radius time curve of the bubble -- he has to plug that into his calculation scheme.

When he looks in the literature, he sees different values for R_0, Pa, the important parameters of sonoluminescence, so he cannot calculate a specific spectrum, but whatever values he gets from the literature he can calculate his spectrum to.

For argon he was able to make two calculations from all the data showing a bound for the spectra; that is, he found the smallest R_0, the smallest Pa, that worked, and the largest R_0 and the largest Pa, and he calculated the spectra for those two cases and he got a dashed curve down here and this bigger dashed curve up here. He is very happy that sonoluminescence data are somewhere between those two curves.

He also calculates a xenon calculation. There are lots fewer data for xenon, so he has only one curve. Again, he is happy, not because the two overlap but because it shows an important distinction between the two types of sonoluminescence, the argon, which does not show a hump down to 200 nm, and the xenon, which does show a hump. Willie Moss is very happy that his calculations lie somewhere in that data range.

When you go through that method, you really do not have a mechanism for the light emission. It is just a bunch of calculations, an opacity table that I do not even understand, but he just plugs in values that are given to him by other researchers, and you can calculate the spectrum or the light emitted by the bubble.

He says that it is probably radiative recombination, so you get an electron coming back down and radiating, or Bremsstrahlung occurring, but that is not actually in his code, so it really does not tell you what the mechanism is.

[Transparency 45]

Another group of researchers has done analytical calculations that are much more simple than having to use the Lasnix [phonetic] codes, the nuclear codes at Lawrence Livermore. Here they make assumptions for what the radiation mechanism is and then they go on from there.

They assume there is thermal Bremsstrahlung. You have electrons near ions and there are neutral atoms giving off light and you have radiative recombination. If you work through the mathematics, you get a power spectrum with a couple of terms. One is the Plank blackbody formula and then this other stuff that incorporates the opacity, again, which is very similar to what Willie Moss showed in the other viewgraph. He showed two terms, one, Plank body (which is shown right here), and then this other term with the opacity (the optically thin part).

I was asking myself this morning what the heck does Kappa actually mean. What do I mean when I say photon absorption? I wrote down something really quick -- I think I am right here. Photon absorption is just the reverse process of photon emission. You can have free-free interactions. You can have bound-free interactions, and you can have bound-bound interactions.

A bound-bound interaction is just an excited state interacting with a lower state and you are getting some known emission, some known energy. That will generate band emissions or line emissions. If you have bound-free or free-free interactions, those drive a continuum but they also are involved in the photon absorption.

MR. GILBERT: Would you say again what free and bound refer to?

DR. MATULA: Sorry. If you have an electron, in one of its shells it is bound. If it is in an excited state and it comes to a lower state or if the molecule is vibrationally excited and it relaxes to another state, that is a bound-bound transition. If an electron comes from the free space, whizzing around, and it relaxes to a shell of one of the atoms, that goes from a free state to a bound state. Because the electron can start off with any energy and it ends up at a specific energy, that can generate a continuum emission.

MR. GILBERT: So what do you mean by free-free interactions?

DR. MATULA: That is a good question. Normally I say it is a good question, because I can answer it, but this one is not so easy to answer.A free-free interaction is when you have electrons and ions interacting with each other and what can happen is, in the photon emission case, an electron can interact with an ion, collide with it and emit a photon, or it can inelastically collide with it and absorb a photon. That is what they mean by free-free.

MR. GILBERT: But they do not mean single electrons, because my remembrance is that you cannot conserve momentum and energy with a free electron, so it cannot absorb a photon completely.

DR. MATULA: Really?

MR. GILBERT: Yes, I think so. It has to be interacting with something else.

DR. MATULA: They say that it is an interaction between free electrons and ions, or neutrals, so can it work that way?

[Simultaneous discussion]

DR. MATULA: These are scattering, they are inelastic collisions.

MR. GILBERT: It does not just absorb the photon and keep it. It re-emits it.

DR. MATULA: It can re-emit it, but can't the ion change its state?

MR. GILBERT: The ion can.

DR. MATULA: And then you can get nonradiative relaxations after that?

MR. GILBERT: Right, by individual free electrons. I will check it and see.

DR. MATULA: There are two different models that I am working with here. One is the Willie Moss model, which is qualitative, because they use Lasnix codes and all the calculations are done with a lot of mathematics that I do not understand. Basically, what is occurring here is you have this opacity as a function of temperature that is different for xenon and argon.

Xenon turns on at a lower temperature, as I mentioned earlier, and so it radiates over a larger volume and it looks hotter than argon. That is how, qualitatively, you look at the Moss model. It generates spectra that do not match the data exactly but it has the same relationship;

argon spectra are a continuum that increases into the ultraviolet, xenon spectra have bumps in them.

In the Lohse model, they actually calculate what the absorbance is, this Kappa. They calculate it assuming hydrogen-like atoms -- which is not really occurring here, you have argon or xenon -- but they make assumptions. Then they argue, by pointing to other references, that it is close enough.

What they get is an absorbance that depends on the ion energy and 2 times K_b Boltzmann's constant times temperature. So Kappa here depends on the ion energy, which is different for xenon and argon; xenon has a lot lower ion energy than argon.

[Transparency 46]

I showed you what the spectrum looked like when you looked at the Willie Moss method. Here is what the spectrum looks like when you look at the Lohse method, and the Hilgenfeldt method. Here is the power, watts per nanometers, as a function of wavelength from 200 to 800 nm. This is calculated for xenon, and this is calculated for argon.

For xenon it matches the bump that you see experimentally. Unfortunately, for argon it does not. It shows a bump for argon as well. Lohse uses a very simple ideal gas type of model to model his emission. Because of that, I think you are not going to ever get perfect matching with data. In fact, I would claim this line right here tells me that his emission mechanism is absolutely wrong.

Of course, if he then takes his emission mechanism and calculates the full width at half-max, the pulse width, has a function of the light emission intensity and he looks at all the data shown in black -- here and here -- and then he calculates that same pulse width, his calculation fits very well.

In one case you can use a relatively simple ideal gas law, taking into account the hard core of the van der Waals hard core, which you need to have, a simple energy mechanism, and calculate full width at half-max that are on the order of 50 to 200 psec, which match the data, and you can calculate spectrum, but every once in a while you will get something that does not work out, like this.

Willie Moss' theory, on the other hand, has the same trends everywhere. The argon spectrum continues up, the xenon spectrum turns over, and Willie Moss also shows the trends here that match data.

One other difference between the two models is in the shape of the pulse. This is the shape of the pulse that we saw earlier. It has an asymmetry to it, it is longer in the tail. Moss does show a very similar looking tail and Lohse's tail is much more spread out.

I do not want to say that that simple model is wrong. What I want to say is that it is a very simplified model. It is an ideal gas interacting. You are assuming hydrogen atoms and you are getting absorption and you are getting photon emissions by very simple models. It is not going to predict everything. It does a very good job of predicting some features like some parts of the spectrum and the pulse widths. It is also very good because you can do hundreds of calculations in a single day, it is very simple.

Willie Moss' model requires coupling a Rayleigh-Plesset-like equation to this nuclear fusion equation, the Lasnix code. On the other hand, he seems to match all the data I have seen out there so far.

MS. HIGHTOWER: If his method is assuming hydrogen atoms, then how can he have a difference between xenon and argon when he calculates stuff?

DR. MATULA: Actually, there are two parts to it. There are the free-bound transitions that occur that he assumes are hydrogen-like transitions -- I forget offhand, actually. I was thinking the free-free was the xenon. He assumes the hydrogen atom is what generates the free-bound transition and his absorbance is calculated through free-free transitions, I think, but that does not make sense, because, to me, still it is an ionization effect that comes into play. I do not know.

So Willie Moss calculates this Lasnix code, generates calculations that seem to match data. The Lohse code can be done a lot faster. It matches some of the data but does not match all of the data, so there is something missing in it.

There is actually another calculation that is very important that does not go into the detail of Willie Moss' nuclear code but expands far beyond Lohse's simple adiabatic model and that is done by Andrew Szeri at Berkeley.

[Transparency 48]

Andrew Szeri takes the gas-diffusion equation, takes the Rayleigh-Plesset equation, does not make assumptions, just plugs it all into his powerful computer and starts cranking out data. What he finds is something very interesting.

Remember those curves I showed you for gas diffusion equilibria? We assumed that no gas is diffusing over one cycle, we just look over long time frames to see if the bubble is stable or not? If you actually look in one acoustic cycle, you can see gas diffusion.

This graph shows what actually happens if you mix argon and helium together. As the bubble collapses, you get a separation or a segregation of the two species. One of the species, and I believe it is helium, will migrate out to the wall and the other species migrates into the center of the bubble, so you actually get segregation of the two noble gas species.

These are brand new data so there are no real experiments that have been done to test whether this hypothesis is true or not. One can envision that if it is true, if this is argon inside here, it should be the one that is giving off the light, not the helium, and you should be able to tell whether it is argon or helium inside the bubble when you mix the two, but it is an interesting calculation to show that there is segregation occurring inside the bubble.

[return to Transparency 47]

He actually includes chemical reactions going on inside the bubble. Chemical reactions can be a pain. This reaction scheme was done by Kamath and Prosperetti several years ago. They wanted to see how much energy was removed from the system by chemical reactions, endothermic reactions, so they looked at water and they looked at some of the reaction schemes that water undergoes.

They have right constants and they can sum up the number of reactions going in one direction that are endothermic and the number of reactions that are exothermic, take the energy difference, plus it into their calculations to see if chemistry is affecting the amount of light being emitted. There are 19 mechanisms here.

There is another guy, named Yasui, who I saw, in his paper, had 63 different schemes -- that is a lot -- to include chemistry.

[Transparency 49]

Chemistry is important, as we have seen before, but what appears to be a very important piece of the puzzle is water vapor. Water vapor, as we have always claimed before, cushions the collapse. If you increase the temperature of water and you generate sonoluminescence, the bubble is much dimmer. If you decrease the temperature, Hiller showed that you can get an order of magnitude or more increase in the sonoluminescence intensity. So we have always said it must be the water vapor cushioning the collapse somehow.

Andrew Szeri shows that in some sense of the word that is indeed what is occurring. He is plotting here the mole fraction, the number of moles of water vapor, divided by the total number of moles inside the system as a function of radial position and he is doing this in a Lagrangian scheme, so basically these are particle tracers.

Here is a bubble 833 nsec before its collapse. It extends out to 23 microns. As it collapses it gets smaller and smaller to this size -- that is what this graph shows, the end points, the radial position.

This is the distribution that he is calculating for water vapor inside the bubble. It is almost uniform through the bubble, but as the bubble collapses, water vapor will condense to the outside until a certain point is reached when the bubble's velocity is so fast you cannot condense water vapor any more and it traps the water vapor.

I am not exactly sure why, but it also looks as if you get this distribution inside the bubble. A lot, of course, stays on the walls and a lot less is actually in the middle of the bubble.

[Transparency 50]

If you include the endothermic reactions that occur when water molecules are torn apart, you get a very interesting result. Szeri plots the maximum temperature that he gets from his bubble as a function of this expansion ratio, the maximum bubble size to the calculated minimum bubble size.

If you do not assume water vapor, if you leave water vapor out of the equation, as the bubble expands to bigger and bigger sizes you just get higher and higher temperatures. If you include water vapor, which is an energy-sucking-up mechanism, you get this green curve and you actually saturate in terms of the maximum temperature you can achieve by single-bubble sonoluminescence. According to his calculations, I cannot just continue to drive this bubble harder and harder; it is going to saturate.

If you include some other chemical reactions you get some more endothermic reactions and more energy sucked out of the system.

[Transparency 51]

That was really bad news for me, because I told people that we had a lithotriptor, and Larry showed it earlier today, that if you take a single bubble and generate a bubble in a lithotriptor pulse, you can expand it to millimeter scales (the calculation he showed earlier), and when that millimeter-size bubble collapses to microns it should emit a lot of light.

In fact, this is a measured radius time curve from a bubble in a lithotriptor pulse. I am proud of this curve, because nobody had ever measured this before. Although people were pretty sure that a bubble would undergo this type of expansion and collapse, we were able to actually scatter light off of a single bubble and measure its growth, collapse, and rebound.

Look at the growth here. It is starting off at about 80 microns and it is growing to over a millimeter in size. If you make just a simple assumption of adiabatic collapse over this time frame, you would get, probably, hundreds of thousands of Kelvin, very hot, but you actually get less sonoluminescence from this than you do in single-bubble sonoluminescence.

According to Andrew Szeri, the reason is simple: You are trapping more water vapor and water vapor is removing energy from the system.

Another thing regarding this particular viewgraph that Larry asked me to talk about is the fact that you also get sonoluminescence when you first compress the bubble, when you take a radial bubble and compress it with a shock wave, so we actually see sonoluminescence at the very beginning and at the very end, but in both cases we do not see as much as we see in single-bubble sonoluminescence.

DR. ATCHLEY: [Inaudible]

DR. MATULA: That brings up an interesting question. How do you take a container of water, stick it into a water bath and generate a standing wave, because your boundary conditions are all shot to hell. The answer is you have a lithotriptor pulse that is being focused to the center, so we actually make our sonoluminescence rig with an air gap everywhere except where the shock wave is going to come through. That is the only region of the cell that sees the water on both sides.

DR. ATCHLEY: It does not matter what phase you are at?

DR. MATULA: Oh, yes, it matters a lot. This is a really nice curve, obviously, or I would not have brought it.

If you hit the bubble during the growth phase, as Larry showed earlier, the bubble will just collapse immediately and probably shatter, and we see that. If you accidentally time the system perfectly so that the shock wave hits the bubble as it is collapsing, we should be able to get much more light emission. We have not timed it well enough yet to get to that, but we are trying.

What I was hoping to do was just the opposite: Instead of worrying about timing the shock wave to hit the bubble during the collapse, I would just use the negative portion of that shock wave and make the bubble expand really big until Andrew told me that water vapor was going to kill me.

Of course, the solution is not to use water. We are going to find a low-water-vapor solution and try it in that.

If you believe what I said about the mechanism for sonoluminescence, that it is thermal, that we can match the data, sonoluminescence is known, we understand it -- a least some aspects of it, or a lot of aspects of it -- where do we go from here?

In fact, not being a very bright guy, but knowing a little bit, I can predict the future of sonoluminescence, and I do that here.

[Transparency 52] (laughter)

Number of publications, starting with the discovery of sonoluminescence -- that is my R_0 -- maximum number of publications. You guys should feel sorry for me: Look at how many publications I had to read last year. A year or so from now there should be a collapse in the publication record. If anybody can tell me why it goes negative, I would appreciate it.

But what is the future of sonoluminescence? There are several interesting aspects. I told you that if you drive the bubble hard enough it will reach what we call the extension threshold and disintegrate. People have calculated that you get surface modes on the bubble by assuming spherical harmonics, but they have never told me what the mechanism for those spherical harmonics is.

One possible mechanism is the buoyancy force. I show that in this viewgraph here.

[Transparency 53]

Let's look at the sound field in our cell, where this is top, this is bottom, for two cycles -- here and here. The dashed line is the gradient of the pressure. During the negative portion of the sound field the bubble grows, it expands. You can qualitatively look at this as seeing that the pressure above the bubble is less negative than the pressure below the bubble, less negative, positive, pushes the bubble down.

At 180° out of phase you have the pressure compressing the bubble. Now the pressure below the bubble is larger than the pressure above the bubble and you have a force in the opposite direction pushing you up, but because the bubble is smaller, the force is smaller, so over an average of an acoustic cycle the net force is to hold the bubble at this pressure antinode -- right here.

The interesting thing is that you have a periodic force on this bubble and that periodic force should translate into a periodic translation of the bubble.

[Transparency 55]

Skipping the mathematics, let's look at what happens if we calculate what a bubble should do in the sound field. Again, this is the sound field, and when it turns positive, you have this demarcation line going from a force toward the antinode to, on the other side, a force away from the antinode.

Below I have the instantaneous force, the acoustic radiation force. For a bubble driven calculated at 1.3 atm, most of the time the bubble is in this region where the force is toward the antinode. You see on the instantaneous force it has mostly a negative component.

If you drive the bubble harder, the bubble will expand for a longer time period and collapse later in the acoustic cycle. Now the bubble spends some time of its motion, where its volume is large, in a region where the force is away from the antinode and you actually get this positive component of the acoustic radiation force popping up.

If you could drive this bubble at 1.7 atm, which you cannot, you would see that positive component growing even larger, so I use this as a qualitative argument that the bubble is undergoing translational motion. It can develop instabilities in the shape, it can die out, it can be extinguished.

[Transparency 56]

This viewgraph shows possible asymmetries in the bubble behavior. This is from Lafont's group in Belgium. They had a single bubble giving off light, and that is shown in A, B, and C, down here, and they very carefully stuck a very small needle with some dye and injected a small amount of dye above the bubble to see what would happen.

The dye came down closer to the bubble and when it reached very near the bubble it started spreading out and, in fact, went back up, not around the bubble, but it went back up. If they injected the dye next to the bubble, it did the same thing, it went back up, vertically upward. The reason they injected it next to the bubble because the first time you would have set off a shock wave from the bubble that is pushing it back up, but now they are injecting it to the side and the shock wave should push it off to the side.In all cases they found that the dye went vertically upward above the bubble.

In a second experiment they let the dye go down to where the bubble was and then they generated a sonoluminescence bubble that was surrounded by this dye. Instead of the dye spreading out in all directions, it all went up above the bubble.

[Transparency 57]

These experiments motivated Michael Longuet-Higgins, who is in San Diego, to calculate what would happen if you had a radially oscillating and translating bubble. What he calculated was that you would get this dipole field of fluid motion -- and I do not remember which way is up and down in this particular picture.

This is not for a sonoluminescence bubble, it is for a linear oscillating bubble. Nevertheless, it is an interesting calculation that shows that you do get this asymmetry. I am arguing that you have this periodic force on the bubble pushing it up and down and there are some other data that show there is some asymmetry going on in the system and we want to get rid of the asymmetry. The best way to get rid of asymmetry is to get rid of its source.

[Transparency 58]

To me, the biggest source of asymmetry in this system is buoyancy. So we get rid of buoyancy by going on the "vomit comet," and I think you all know why it is called the vomit comet. It is a roller coaster ride. The airplane goes up into a 45° nose-high position, generates about 1.8 G's of acceleration, then it turns over and, as it is turning over, you are basically falling and the plane is going to fall with you for about 20 seconds or so, then it pulls out before it hits the ground.

What this allows me to do is an experiment in micro-gravity -- it is not really zero gravity, it is micro-gravity. It allows me to try to remove some of this asymmetry that I am arguing is occurring inside the cell.

[Transparency 59]

This particular experimental setup is relatively simple to do. You just use laboratory equipment, a function generator, power amplifier, power supply for the PMT. Everything has to be done in a light-tight box because of the lights in the system and we have a video camera monitoring what is happening.

Obviously, it is a lot of fun to go up in the vomit comet, it is a blast, but not everybody has a lot of fun.

Here is my experimental apparatus right here when I sent one of my students, and I am looking for him and I do not see him anywhere around here. (laughter)

He is not enjoying himself. He had a pretty hard time. Fortunately, the experiment ran without him. I knew him well enough to be able to computer-control the whole experiment. All he had to do was start the equipment and the computer, using a solid-state relay, generate a bubble, monitor its stability, track its motion, run through the experiments that I wanted it to run through, and all he had to do was shut it off at the end.

The good thing about this is I do not have to worry about my students, whether they are sick or not. The bad thing is the students are not going to work, they are going to sit there and play the whole time, and they do.

[Transparency 60]

What happens with the data now? That is the important part. The red line shows the gravitational acceleration. You go into the parabola, you go from about 1.8 G's down to micro-gravity for about 20 seconds or so. You come back out of the parabola and you just keep on going through this parabola roller coaster ride.

The blue line shows the light intensity. What I am doing here is I am holding everything absolutely constant, I think, except for G. As a physicist I am just varying G. I am not driving it any differently and I am seeing a large increase in light emission -- I am up to a 40%, sometimes 50% increase in light emission when it goes from 2G to micro-G.

You have to be careful when you go on this airplane, because things other than G are changing. The shape of the parabola, you have to worry about whether the orientation of your system is causing an effect (it turns out that is not a problem). The weather conditions, the pilot's ability to hold micro-gravity. If you look at these micro-gravity data right here it is actually very poor, it is not a very good day, I think, that we had in micro-gravity, a lot of jitter in the G.

The biggest effect is the fact that your bubble is actually a few centimeters below the surface of your water and there is actually this column of water above your bubble, hydrostatic pressure, <u>rho</u> GH above G. As soon as you remove G, that hydrostatic ambient pressure changes and, in fact, if we removed the hydrostatic pressure or we changed the ambient pressure by the hydrostatic pressure by <u>rho</u> GH by some value, we can see changes in the light-emission intensity.

Part of this can be resolved by just saying the hydrostatic pressure is changing in a liquid, the ambient pressure is changing. When you go to micro-gravity, you are decreasing the total ambient pressure on the bubble and you get increased light emission and when you go to hyper-gravity you are increasing the ambient pressure and your light emission goes down.

That accounts for about 5% of the light emission intensity; we have done the experiment in the lab to make sure that that is accounted for. So what we are seeing is sometimes a 40% change in light-emission intensity. I think there is the possibility that buoyancy does play a role in single-bubble sonoluminescence and I am trying to convince NASA to send me up on the mission to Mars or some place to do this experiment.

[Transparency 61]

I want to finish by thanking all the students and other people who have helped me and given me advice for this work. Thank you.

DR. GARRETT: Does the bubble go into a different part of the standing wavefield as you change gravity, because at one point it is working against gravity and at one point it is not working against gravity? Even though the amplitude of the drive as measured at any given point is constant, the amplitude that the bubble sees depends on how close you are to the center of the sphere or whatever you are using.

DR. MATULA: That is, in fact, occurring. When you look at the bubble -- let's look at it theoretically and experimentally -- experimentally, when you go into micro-gravity the bubble shifts downward. The reason is, when you are in the laboratory the bubble is not at the pressure antinode, it is above the pressure antinode because of buoyancy, the average force is zero.

When you remove buoyancy, the bubble drifts back down to the pressure antinode. How much of a drift occurs? We are using a spherical cell, so we calculated with a spherical Bessel function the amount of pressure change that the bubble would see when it did this change. What we saw was that the change in pressure amounted to a change in intensity of the light by about 1%, so smaller than the hydrostatic pressure but definitely there and definitely in the same direction, so that you actually have to add the 5% correction with the 1% correction.

The amount that it moves is very small. It does not move as much as you might think. Maybe that is the confusion. I forget the numbers offhand for how much it moves.

MR. PETCULESCU: [Inaudible]

DR. MATULA: You asked, I think, two questions. People wondered whether this was a laser and I have heard -- they give it some other weird names. I looked at one of those calculations and it just did not make any sense to me at all, because the guy was talking about -- I do not remember offhand, now that I think about it, but I remember it did not make very much sense, so lasing does not appear to me to be a practical effect of sonoluminescence.

Coherence: I think Putterman tried to measure that and I do not think he saw -- I do not think he has ever published it but I think I remember talking to him back in the early 1990s and there was nothing there, it was not coherent radiation. I am pretty sure he tried that measurement and did not measure anything, but my memory is failing me.

DR. KEOLIAN: As I remember, he measured the angular distribution --

DR. MATULA: He did that measurement, he measured the angular distribution. Several people have done that. If I remember correctly, he measured the angular distribution to see if the bubble was asymmetrical and argued that there is some asymmetry in the light intensity as a function of angle and he was able to argue that the bubble was slightly flattened at emission, it had some ecentricity to it. I remember that publication.

DR. ATCHLEY: He looked at polarization as well.

DR. MATULA: He looked at polarization but I do not think he published that and I do not think he saw any polarization effects.

I do know of a group, along the same lines as looking at angular distributions -- nobody actually knows what the emission spot size is, what is the volume size of this emitter. I have been told that in astronomy you measure the size of the emitter, the size of the star, by doing something called the Heinberry-Brown-Twis experiment, where you have a couple of detectors at some angular distance and, through magic or quantum mechanics, either one, you have wavefunction overlaps that occur and you actually can measure over a small angular distance the size of the emitting star.

Somebody tried to do the same thing with single-bubble sonoluminescence, they tried to use angular measurements to measure the size of the light-emitting region, but they were unsuccessful in that.

DR. WILEN: I heard early on that one of the mysteries was that jitter in the light emission was much slower, orders of magnitude slower, than the jitter in the driving frequencies?

DR. MATULA: Yes, that one viewgraph I showed earlier where you have the flash occurring every acoustic cycle with about a 100-psec full-width half-max, if I remember correctly, they said that their instrumentation was not that good at jitter, if they had a higher jitter in their instrumentation, how could they generate flashes that have smaller jitter? It does not make sense.

The only argument I have heard to counter that is the fact that if you take into account the Q of the cell, high-Q cells, the jitter should decrease as 1/Q. I have heard that argument, I am not sure it has been proved, but it makes sense to me.

DR. GARRETT: If the Q is, say, 10,000 -- what you are saying is the jitter in the oscillator is higher than the jitter in the flash -- but if the Q is, say, 1000, that means that each time the oscillator is putting in only 1/1000 of the amplitude, so if it is off a little bit it is not going to amount to much, because you have established a resonance where you are just kicking it at slightly the wrong time with a much lower --

DR. MATULA: That is a good way to explain it.

MR. GLADDEN: In the RT curves, is radius a function of time for the bubble? When you roll off and you have that initial collapse and, actually, for each subsequent collapse, I am sure someone must have calculated what kind of acceleration you are talking about right at that cusp there where there is that sort of hard sphere limit. I am just wondering what kinds of magnitudes there are.

DR. MATULA: People have calculated it but, better yet, people have measured it. People have measured down to the limits of resolution of their system to bubble collapse and the rebound. Putterman did this with a very fast femtosecond laser, strobing it, and a group in Germany did it using light scattering into a streak camera.

I believe the number was astronomical. It was like 10^{10} G's. It was just an astronomical number. That was what made sonoluminescence a very controversial subject for a lot of years. It appears to be an exotic regime, such fantastical numbers are associated with it that people started putting out all sorts of theories. At one time I counted 14 different theories for sonoluminescence that extended from everything from squeezing a bubble, it gets hot, to electrical discharges, because bubbles are known to carry electrical charges on them, to quantum vacuum radiation. We had seen a talk earlier in the week on the Kasimir effect. You have two forces on parallel plates and you saw the acoustic analogy to that, which is really cool.

Schringer postulated that sonoluminescence could be due to a dynamical Kasimir effect, where you had the bubble moving and then real photons being created out of virtual photons and those being emitted. It took about seven or eight years over all the calculations. That is a possibility but the number of photons being emitted is so small as to be minuscule and not really part of what we see, probably.

Other theories? There were theories that involved the bubble being unstable and I think Larry Crum showed a jet going through the liquid -- that is a famous picture of Larry Crum's -- and if the jet goes through the liquid it smacks the liquid below the bubble, it can fracture the liquid, and you get something called fractoluminescence; somebody was arguing that a while back, that that should generate light emission that depends on the gas that is dissolved in the liquid.

Chemiluminescence: People have been arguing for years that it is just chemistry, it is not really getting hot, it is free-bound transitions, free-free transitions, bound-bound transitions, just normal chemistry. In fact, it is kind of like that but it is just a little bit hotter, it appears to be a little bit hotter.

Most people believe now that it is just a thermal mechanism. Chemistry occurs from a thermal mechanism as well, of course. It is not very hot. Most people believe now that it is 10,000 to 20,000 Kelvin, which is still hot, much hotter than the sun.

That reminds me of that question I asked at the very beginning. If the sun is 5000 Kelvin and it is yellow, and multi-bubble sonoluminescence is 5000 Kelvin, why is it blue?

PARTICIPANT: It is not blackbody, so --

DR. MATULA: But the sun is, right. That is my explanation for it.

MS. HIGHTOWER: But the sun is not yellow in space.

MR. PORTER: Yes, that is what I have heard. After it goes through the atmosphere it turns yellow.

DR. MATULA: But the peak in the spectrum at 5000 Kelvin, isn't that yellow?

[Simultaneous discussion]

DR. MATULA: There are two parts to the outer layer. One is the corona, which is 5000 Kelvin. That is the one I am talking about.

MR. GILBERT: But it is much hotter in the interior.

DR. BASS: In the photosphere what you see is 5800. The corona is much hotter. You never see it, so you do not worry about it.

DR. MATULA: Yes, that is right.

Thank you.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE **30 Sep 01**	3. REPORT DYPE AND DATES COVERED **Final 22 June 98 – 30 Sep 01**

4. TITLE AND SUBTITLE
Proceedings of the 2000 Physical Acoustics Summer School: Volume 1. Transcripts

5. FUNDING NUMBERS
PE 61153N
G N00014-98-1-0044

6. AUTHOR(S)
Henry E. Bass and Elizabeth A. Furr

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Jamie L. Whitten National Center for Physical Acoustics
The University of Mississippi
University, MS 38677

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
ONR 331
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Volume I of the Proceedings of the 2000 Physical Acoustics Summer School contains verbatim transcriptions of the lectures given at the summer school. Volume II of these proceedings contains copies of the transparencies used by the lecturers and Volume III contains background materials that were sent to student and discussion leader participants prior to the summer school.

14. SUBJECT TERMS **Physical Acoustics, Nonlinear Acoustics, Acoustics Demonstrations, Noise and Sensors, Resonant Ultrasound Spectroscopy, Atmospheric and Meteorological Acoustics, Thermoacoustics, Medical Applications of Acoustics, Sonoluminescence**

15. NUMBER OF PAGES
518

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-1
298-102