

2



# Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

RECEIVED  
LAWRENCE  
BERKELEY LABORATORY

OCT 19 1983

LIBRARY AND  
DOCUMENTS SECTION

QUANTUM EFFECTS IN THE DYNAMICS OF BIOLOGICAL SYSTEMS

W.S. Bialek  
(Ph.D. Thesis)

September 1983

**TWO-WEEK LOAN COPY**

*This is a Library Circulating Copy  
which may be borrowed for two weeks.  
For a personal retention copy, call  
Tech. Info. Division, Ext. 6782.*

**Donner Laboratory**

**Biology &  
Medicine  
Division**

LBL-16642  
2

## **DISCLAIMER**

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

QUANTUM EFFECTS IN THE DYNAMICS OF BIOLOGICAL SYSTEMS

William Samuel Bialek  
(Ph.D. Thesis)

Biology and Medicine Division  
Lawrence Berkeley Laboratory

and

Department of Biophysics and Medical Physics  
University of California  
Berkeley, CA 94720

September 1983

# Quantum Effects in the Dynamics of Biological Systems

William Samuel Bialek

*Department of Biophysics and Medical Physics, and  
Biology and Medicine Division, Lawrence Berkeley Laboratory  
University of California, Berkeley  
Berkeley, California 94720*

## ABSTRACT

The performance of biological sensory systems is shown to reach the quantum limits to measurement, this being true in spite of the high levels of thermal noise associated with operation at physiological temperatures. Theoretical issues associated with quantum-limited measurement at high temperatures are addressed and strategies for such measurements which make use of active filtering are formulated. Experimental and theoretical evidence supporting the existence of active filters in the sensors of the inner ear is discussed.

Quantum-limited measurement requires that the internal dynamics of the primary sensory amplifier have a quantum mechanical "phase-memory" comparable to the time scale of the measurement, on the order of milliseconds for biological sensors. This millisecond coherence time invalidates conventional (and semi-classical) descriptions of biological systems in terms of chemical kinetics, and demands an analysis of quantum effects in molecular dynamics.

Simple model Hamiltonians are formulated which seem to describe the dynamics of biological molecules and field-theoretic techniques are developed to solve these

Hamiltonians in a variety of parameter regimes. Conditions for non-trivial quantum effects in these models are compared with experiments on the primary events of photosynthesis and the low-temperature behavior of heme proteins. Predicted quantum effects are observed and independent observations support the assignment of parameters outside the regime where semi-classical approximations are valid.

The same model Hamiltonians predict that molecular vibrations exhibit a super-radiant instability in polymers where each subunit catalyzes a chemical reaction of sufficient speed. This instability results in macroscopic quantum coherence of the sort required to account for quantum-limited measurement in the sensory systems. Such "phonon super-radiance" provides a natural mechanism of amplification and active filtering, several features of which are in accord with experiments in the sensory receptors of the inner ear.

In each case considered, theoretical analysis demonstrates that quantum effects in biology are significant if not dominant, and this leads to fundamentally new interpretations of several biological processes. Detailed experiments are proposed which will provide quantitative tests of these new interpretations.

## Preface

*...One of the difficulties of scientific research is that it is impossible to make progress without clear understanding, yet this understanding comes only from the work itself...*

*A.B. Migdal, 1975*

All of the support for the work contained in this thesis has come from the State of California and the federal government. I have had the good fortune to be born in a state which had, at least for a time, seriously committed itself to the education of its youth. In exchange for my high school diploma, itself obtained from a public school, I was entitled to a nearly free education at one of the world's finest Universities. At this time, our State and its University finds itself in the midst of financial crises, as does the nation as a whole. In response to these crises many of the programs which have aided my education have been reduced and the nearly free education which I received has all but been abolished. It is therefore with much concern that I write these acknowledgements, concern that I shall be among the last generation of students to be so generously supported.

The work reported in this thesis was supported by the National Science Foundation Pre-Doctoral Fellowship Program, by the National Science Foundation Biophysics Program (Grant No. PCM 78-22245), and by the Office of Basic Energy Sciences,

Office of Energy Research, U.S. Department of Energy, under Contract No. DE-AC03-76SF00098. The opinions expressed herein are those of the author and not necessarily those of the supporting agencies.

Some of the results presented here were obtained in collaboration with Bob Goldstein and Allan Schweitzer, and most of the ideas were discussed with them at one time or another. Both Bob and I were carefully watched over by Professor Alan Bearden, who magically knew when to advise us and when to leave us be. I shall always remember Alan's consternation upon realizing that Bob and I spent most of our time in the office chatting and how his fears were relieved when we presented our first theoretical results to him; not long thereafter we acquired a larger office.

One of the pleasures of Berkeley is that one learns as much from one's fellow students as one does from the faculty. In particular, Alan seems to attract an unusual collection of students, which makes life fun for all of us; while I have been here I've been educated by and enjoyed the good humor of Mark Crowder, Sharon Hibdon, Bob Hootkins, Rocky Nevin, Jim Potter, and of course Bob G.. From the first day I arrived at Berkeley I have shared all my interests with Dave Chernoff and Dan Seligson, and their patient and critical ears have done much to help my own thinking.

The life of a student in this department is made easier by Dierdre Anderson, Kathleen Campos, Jean DeGrassi, and Marilyn Nichols. As an NSF Fellow my existence was greatly simplified by Nancy Plunkett, who solved many of my personal "fiscal crises" by knowing how the University bureaucracy operates.

The traditional ordeal of the qualifying examination was made pleasant by cheerful and enthusiastic examiners: Professors Orlando Alvarez, Marvin Cohen, Anastasio Melis, Jack Owicki, and Paul Richards. Orlando was so enthusiastic that he got stuck reading this thesis, a task he shared with Alan and Professor W. Geoffrey Owen. I am

happy that Geoff did not hold against me our experience of teaching a course together.

Bob G. and I certainly could not have accomplished as much as we did without tremendous assistance from the staffs of the Three C's, the Espresso, and Aki's. The opening of Pain et Chocolat accelerated our progress while the demise of the China Garden--and with it the ideal pot sticker--was a serious setback.

I have numerous specific scientific debts: Professor E.R. Lewis and his collaborators shared their data with me in advance of publication, as did Dr. Michael Wilson (now at University College, London), while Ken Krieg (Berkeley) took a potentially large computing problem off my hands. Dr. Carlton Caves (Cal Tech) gave me several useful comments on the issues of Sect. II.B and told me about his related and unpublished work; similarly with Professor John Hopfield (Cal Tech) *re* Sect. III.B. Finally, Professor George Zweig (Cal Tech and Los Alamos) spent several long afternoons in Miami forcing me to clarify my ideas about noise in the sensory systems, making much of the further development considerably easier.

The completion of a thesis involves much more than science. Similarly, Alan has served in these last few years as much more than a scientific adviser. My parents, who nurtured my interest in science and education in general have had to watch as the pursuit of these interests forces me to spend more time at work and less time with them. Charlotte agreed to live with me and married me at what will be the most difficult time of our lives together, as she provided emotional stability, a second income, and in the end even the beautiful figures which grace these pages.

## Table of Contents

<b>Preface</b> .....	ii
<b>I. Introduction</b> .....	1
<b>II. Macroscopic quantum effects in biology: The evidence</b> .....	9
<b>A. Photon counting in vision</b> .....	9
<i>1. Lorentz' prediction and the Rose--de Vries law</i> .....	9
<i>2. Photon statistics and the frequency of seeing</i> .....	10
<i>3. Single photon signals from individual receptor cells</i> .....	14
<i>4. Models for quantum counting</i> .....	18
<b>B. Modern views of macroscopic quantum mechanics</b> .....	24
<i>1. Schrödinger's equation for a ton of brass</i> .....	24
<i>2. When is the Heisenberg uncertainty principle important?</i> .....	25
<i>3. Bridging the gap between quantum and classical regimes: Amplifiers</i> .....	30
<i>4. Conclusions: Quantum-limited measurement at finite temperature</i> .....	42
<b>C. Zero-point motion: Mechano-receptors of the inner ear</b> .....	44
<i>1. Can we really hear sub-angstrom motions?</i> .....	44
<i>2. The ear at <math>T = 0</math>: Quantum noise and the threshold of hearing</i> .....	49
<i>3. The ear at finite temperature</i> .....	50
<i>4. Models for a linear amplifier</i> .....	61
<b>D. Some conclusions</b> .....	70

<b>III. Molecular dynamics and macroscopic implications</b> .....	73
A. A general theoretical approach .....	73
1. <i>The experimental background</i> .....	73
2. <i>Model Hamiltonians for biological molecules</i> .....	86
3. <i>Computational methods</i> .....	93
4. <i>Some predicted features of biomolecular dynamics</i> .....	115
B. Resonance in the slow regime: Theory vs. experiment in myoglobin .....	123
1. <i>Vibrational lineshapes and non-exponential kinetics</i> .....	123
2. <i>Direct measures of vibrational lineshape</i> .....	131
3. <i>Vibrational spectra and intramolecular motions</i> .....	140
4. <i>Towards more detailed tests of the theory</i> .....	153
C. Coherence in the picosecond regime .....	156
1. <i>When is an intermediate an intermediate?</i> .....	156
2. <i>Long pulse vs. short pulse experiments</i> .....	168
3. <i>Testing the theory</i> .....	172
4. <i>Some conclusions</i> .....	174
D. Mechanisms for macroscopic quantum behavior in biological systems .....	176
1. <i>Phonon super-radiance</i> .....	176
2. <i>Building the perfect amplifier</i> .....	182
3. <i>Amplifier dynamics and experiments in the inner ear</i> .....	188
<b>IV. Concluding remarks: Where do we stand?</b> .....	200
<b>Appendices</b> .....	209
A. The sensory "threshold" .....	209

B. Further comments on photon counting .....	214
C. Mechanics of the cochlea .....	217
D. Radiationless transitions via the Bloch equations .....	221
E. Radiative interactions in strongly coupled systems .....	229
<b>Notes and references</b> .....	<b>234</b>
Notes and references, Chapter I .....	234
Notes and references, Chapter II .....	236
Notes and references, Chapter III .....	241
Notes and references, Chapter IV .....	247
Notes and references, Appendices .....	249

# Chapter One

## Introduction

*...The physicist is familiar with the fact that the classical laws of physics are modified by the quantum theory, especially at low temperature. There are many instances of this. Life appears to be one of them, a particularly striking one....The living organism seems to be a macroscopic system which in part of its behaviour approaches to that...conduct to which all systems tend, as the temperature approaches the absolute zero and the molecular disorder is removed...*

*E. Schrödinger, 1944<sup>1</sup>*

In the second half of the nineteenth century, the vitalist view of biology began to give way under pressure from developments in the physical sciences. Mayer and Helmholtz demonstrated conservation of energy in living systems,<sup>2,3</sup> Pasteur began to study what we now call the structural chemistry of biological molecules, and Volta and Helmholtz investigated the<sup>4</sup> electrical basis of signal transmission along nerves. The problem of color perception played a major role in the discussions of Maxwell<sup>5</sup> and others<sup>6</sup> concerning the electromagnetic basis of optics, and Helmholtz<sup>7</sup> made use of classical mechanics and the new mathematical results of Fourier to understand the basis of hearing and the perception of harmony.

These remarkable results from the golden age of classical physics still dominate our thinking about the basis of life. Quantitative studies of cellular and molecular

biology are based on thermodynamics and structural chemistry, and these concepts are supplemented by classical electromagnetism, optics and mechanics when they are applied to sensory and neural systems. At the same time, of course, there has been a revolution in the physical sciences with the rise of quantum theory.

The most significant influence of quantum mechanics on biology has been through the rules of chemical bonding. In the 1930's, shortly after the development of quantum theory in its modern form, Slater, Pauling, London and others<sup>8</sup> presented the now well known arguments regarding the quantum mechanical basis of molecular structure and chemical reactivity. Pauling went on to apply these results to the biological problems of catalysis<sup>9</sup> and immunology.<sup>10</sup> His view that molecular structure determines biological function prevails to this day.

The success of Pauling's "structure-function" view of life has led to the belief that quantum mechanics has no implications for biology beyond the rules of chemical bonding. This claim, although widely accepted, has been challenged many times since the inception of quantum theory. At the macroscopic scale, Lorentz<sup>11,12</sup> suggested in 1916 that the quantization of the electromagnetic field may limit the sensitivity of vision, while Bohr<sup>13</sup> argued in 1933 that quantum limits to other biological sensory systems would be recognized once the corresponding physical principles were clarified. At the microscopic scale, a number of workers have proposed that quantum mechanical effects such as resonance<sup>14,15</sup> and tunneling<sup>16</sup> may contribute to or even dominate the interactions among biological macromolecules. In particular, the structure-function view met its first direct challenge in 1938, with a paper by P. Jordan.<sup>14</sup>

Jordan suggested that the quantum mechanical resonance interaction between two biological molecules might be large, and that the specificity of this interaction--which depends on a near-degeneracy of states in the two molecules--could account for the observed specificity of biological processes. This idea was pursued by Jehle,<sup>17,18</sup> who showed that the necessary specificity would occur if the vibrational spectra of the

molecules exhibited well-resolved lines. Pauling and Delbrück<sup>19</sup> responded by asserting that macromolecules in solution could never possess well-resolved spectra, and offered a nearly modern version of the structure-function principle as an alternative explanation of biological specificity.

Modern textbooks<sup>20,21</sup> present molecular structure as the full explanation of specificity in biochemical processes, and no mention is made of quantum mechanical effects as proposed by Jordan and Jehle. In no system, however, has the observed specificity been accounted for *quantitatively* in terms of structural features, and one recent attempt to interpret inter-molecular binding energies in structural terms failed by many orders of magnitude.<sup>22</sup> Thus although the structure-function view of biological specificity dominates modern thinking on the subject, the evidence does not convincingly exclude the possibility of quantum mechanical effects.

The exchange among Jordan, Jehle, and Pauling and Delbrück marks the beginning of a continuing conflict between structural and quantum mechanical interpretations of biological events; this conflict takes many forms. In photosynthesis,<sup>23-26</sup> for example, photon absorption by an "antenna" chlorophyll molecule raises this molecule to an excited electronic state. The energy of this excitation is transferred from molecule to molecule within the antenna until it reaches the specialized molecules of a "reaction center," where it drives an electron transfer reaction; the photon energy is stored in the resulting charge separation and ultimately used to synthesize compounds useful to the plant or bacterium. In pioneering papers, Franck and Teller<sup>27</sup> and J. R. Oppenheimer<sup>28</sup> sought to understand the energy transfer processes in explicitly quantum mechanical terms, while the observation that the electron transfer reactions proceed at very low temperatures with temperature-independent rates<sup>29,30</sup> suggested to many workers<sup>16</sup> that quantum mechanical tunneling contributes significantly to these processes. In spite of these efforts, the dominant concepts remain structural and thermodynamic, not quantum mechanical.

In the structural view, a molecule which participates in an electron transfer reaction is expected to have structural features which allow the rate of this reaction to be controlled by the organism. Crystallographic studies of such molecules<sup>31,32</sup> are interpreted in just this way, with only passing reference to tunneling and other quantum effects in molecular dynamics. Even those papers which focus on tunneling use this quantum effect to derive structural information, such as the width of the barrier to tunneling.<sup>33</sup> Similarly, studies of photosynthetic antennae<sup>34</sup> center on the structural organization of the system--distances and orientations among molecules--rather than the quantum mechanical parameters which characterize this structure and its dynamics.

From these examples it is clear that the question of quantum effects in biology remains unresolved. The consensus of the community, however, is that these effects are at most perturbations to an essentially classical system, so that biological events may be interpreted in terms of molecular structure and the rules of chemical bonding with little regard for quantum mechanics. The clearest challenge to this point of view comes from the possibility that quantum mechanics is *essential* for understanding at least some biological processes.

In a 1944 monograph<sup>1</sup> *What is Life?*, Schrödinger presented a concerted attack on the view of life espoused by a "naive classical physicist." The classical physicist, Schrödinger explained, seeks order in the law of large numbers. The behavior of individual molecules is chaotic, and we can return to determinism only by averaging the behavior of thousands or millions of molecules. Unfortunately, as had been shown convincingly by Delbrück and co-workers, the evidence on this point is against the classical physicist: single-celled organisms carry their most precious possession, the genome, in a single molecule.

Schrödinger pointed out that since classical physics cannot account for the stability of atoms or molecules, the stability of genetic material stored in a single molecule must be viewed as a quantum mechanical phenomenon. Even a semi-classical

approximation to quantum mechanics, such as the Thomas-Fermi method, cannot account for the stability of molecules,<sup>35</sup> so that the stability of the gene as it passes from generation to generation is a quantum effect of considerable subtlety. From this and other examples, Schrödinger concludes that it is only through quantum mechanics that life can be understood in a manner consistent with experimental evidence; his clearest statement is in the passage quoted at the beginning of this Chapter. While Schrödinger's book is renowned for having brought many physicists to biology,<sup>36</sup> not all of these physicists seem to share his enthusiasm for a quantum mechanical view of life. Instead many have embraced Pauling's view, which is a strict reductionist picture of the sciences: biology is based on chemistry, chemistry is based on physics, and there are no routes by which physical--in particular quantum mechanical--principles can directly influence biology.

The question of whether quantum mechanics directly affects the dynamics of biological systems has a very different meaning today than in 1944. Even the simplest of biological molecules is a complex interacting system with several thousand degrees of freedom, and at the time of Schrödinger's writing no such many-body problem had been solved; the issue which had been raised by Bohr regarding quantum limits to sensory systems other than vision had also not been resolved. Since the publication of Schrödinger's book powerful theoretical methods have been developed for describing the quantum mechanics of both microscopic many-body systems<sup>37-39</sup> and macroscopic systems such as sensors and measuring devices.<sup>40-42</sup> Over the same period advances in experimental technique have provided more reliable information on the structure and dynamics of biomolecules<sup>43-45</sup> as well as on the behavior of the individual cells that are responsible for sensory reception.<sup>46-48</sup> I believe that these advances force a reconsideration of the role of quantum mechanics in biology; this dissertation provides such a reconsideration.

We may look for two distinct classes of quantum effect in biology. First is a macroscopic quantum effect, typified by the ability of a sensory system to detect signals near the quantum limits to measurement. Second is a microscopic quantum effect, in which the dynamics of individual biological macromolecules depart from the predictions of a semi-classical theory. Finally, if such quantum effects in fact occur, we must try to understand the macroscopic effects in terms of the microscopic, in the same sense that we understand the macroscopic quantum behavior of a superconducting device in terms of the microscopic behavior of the interacting electrons in the superconductor. These issues are addressed in the following chapters, and the main results are summarized here:

[1] Quantum-limited measurement occurs in several biological sensory systems, including the displacement sensors of the inner ear. Quantum limits to detection are reached in the ear in spite of a seemingly insurmountable level of thermal noise.

[2] In order to reach the quantum limit the receptor cells of the inner ear must possess amplifiers with noise performance approaching the limits<sup>42</sup> imposed by the uncertainty principle. Such "perfect" amplifiers cannot be described by any chemical kinetic model, nor by any quantum mechanical theory in which the random phase approximation is valid--the molecular dynamics of the amplifier must be such that quantum mechanical coherence is preserved for times comparable to the integration time of the detector.

[3] The dynamics of biological molecules may be described by model Hamiltonians similar to those used in several problems of condensed matter theory, albeit in different parameter regimes. Methods are developed for solving these models, which consist of strongly coupled electronic and vibrational degrees of freedom, and conditions for quantum effects on the molecular dynamics are identified.

[4] Two major quantum effects in molecular dynamics--resonant dependence of electronic transition rates on vibrational frequencies and coherent evolution of the

electronic states for times comparable to vibrational relaxation times--are compared with experimental studies on ligand binding in heme proteins and electron transfer in photosynthesis, respectively. Predicted consequences of the quantum effects are observed, and in the case of heme proteins a number of independent experiments are explained by the same model Hamiltonian, strongly supporting the identification of quantum effects in this system.

[5] In systems with long vibrational relaxation times the phonon modes can be pumped by chemical reactions (transitions among electronic states), and this leads to a phonon instability. In the case of a polymer this instability is similar to super-radiant emission of photons by an ensemble of excited atoms,<sup>49</sup> and results in a coherent oscillation of the phonons throughout the polymer. Such macroscopic phase coherence is precisely what we require to understand the existence of quantum-limited measurement (point [2] above), and the phonon instability can be used to construct an amplifier; several qualitative features of such an amplifier correlate quite well with experimental observations on responses of the receptor cells in the inner ear.

Taken together these results provide convincing evidence that, as Schrödinger suggested, at least some phenomena of life can be understood only in terms of quantum mechanics. This is, in spite of all the hints in the earlier literature, a surprising conclusion. For this reason I have sought to make the discussion as concrete as possible and to avoid the more philosophical issues which surround the application of quantum mechanics to biology.<sup>13</sup> Instead I have confined my attention to a small number of systems where theory and experiment may be rigorously compared. In this way it may be seen that quantum mechanical effects are not merely embellishments on a nominally classical system; rather their existence forces essential changes in the way we think about the extant observations. While a focus on specific systems is essential for establishing the existence of quantum effects in biology, it is clear that once such effects are established our thinking about the molecular basis of all biological events

must be re-examined. I hope that the theoretical approaches developed in this work provide a useful starting point for such investigations.

## Chapter Two

### Macroscopic quantum effects in biology: The evidence

*...The most striking features are: first, the curious distribution of the cogs in a many-celled organism....and secondly, the fact that the single cog is not of coarse human make, but is the finest masterpiece ever achieved along the lines of the Lord's quantum mechanics...*

*E. Schrödinger, 1944<sup>1</sup>*

#### A. Photon counting in vision

##### *1. Lorentz' prediction and the Rose--de Vries law*

In 1916, eleven years after Einstein introduced the photon, Lorentz used the existing data on the visual threshold to estimate the minimum number of photons required at the cornea for a human subject to respond. He found that this number is on the order of one hundred or less.<sup>2,3</sup> Realizing that not all of these photons reach the retina, he suggested that the threshold of vision is limited by the quantal nature of light. This suggestion may be the first formulation of a quantum limit to measurement, and it certainly was the first application of quantum mechanics to biology.

The quantization of light can influence the reliability of our perception even at light intensities well above the threshold of vision. Any attempt to distinguish

between two dim lights will be limited by the variations in the number of photons arriving at the retina, in addition to whatever sources of noise are contributed by the eye itself. Thus, if the photon statistics of the light source are Poisson, then an average of  $N$  photons absorbed by the retina is accompanied by a standard deviation of  $N^{1/2}$ . This variance provides a limit to the discriminability of changes in the intensity of the source, and predicts that the threshold for intensity discrimination should grow as the square root of the intensity. This square-root relation was first applied to vision by de Vries<sup>4</sup> in 1943 and Rose<sup>5</sup> in 1948.

The Rose--de Vries law must break down at both very high and very low intensities, as illustrated in Fig. II-1. At very high intensities, fluctuations in the gain of the photoreceptors must dominate any fluctuations inherent in the signal, and will lead to "Weber's law," namely intensity discrimination thresholds proportional to intensity. At the very lowest intensities, the photon statistics variance will fall below the "dark noise" level of the photoreceptor, and the intensity discrimination performance will be limited only by this intensity-independent internal noise. Barlow<sup>6</sup> has summarized the evidence that these three regions, dark-noise limited, photon statistics limited, and Weber's law, are indeed found in psychophysical data from human observers as well as in the firing patterns of single cells in the optic nerve of the cat. These studies strongly suggest that the statistics of photon arrival at the retina limit the reliability of our judgements of light intensity, as Lorentz had predicted.

## *2. Photon statistics and the frequency of seeing*

In a series of psychophysical experiments, published in 1942, Hecht, Schlaer and Pirenne<sup>7</sup> and (independently) van der Velden<sup>8</sup> addressed the issue of photon statistics more directly. They showed that the detection of a weak visual stimulus by a human observer is a probabilistic event, and that the statistics of this event are dominated by

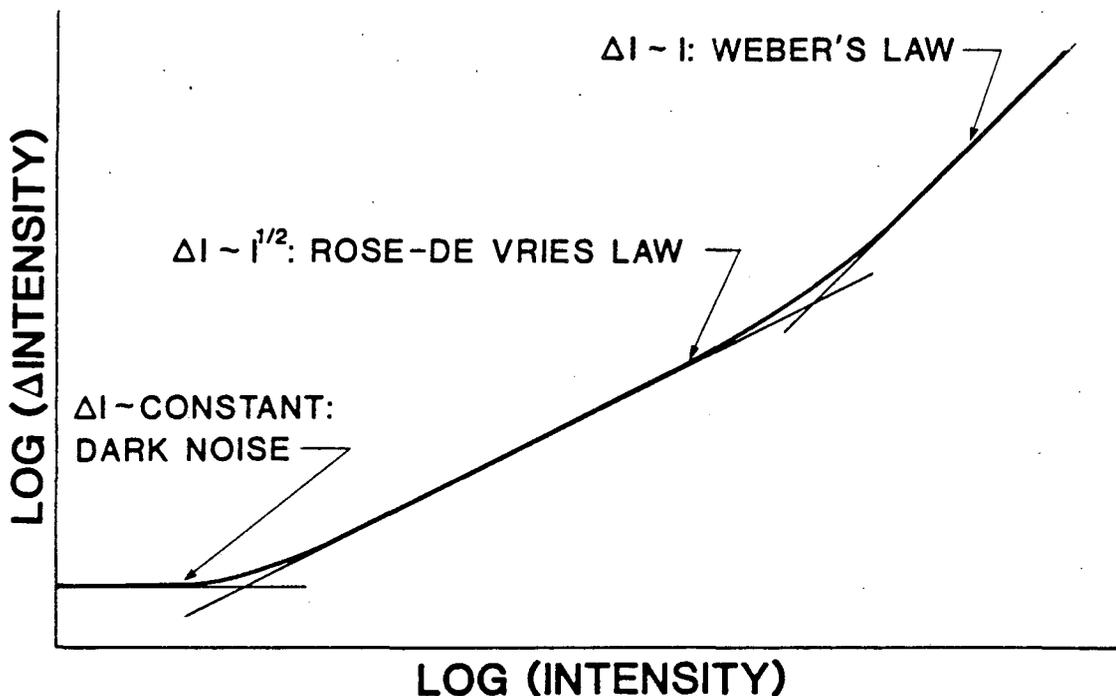


FIGURE II-1 Intensity discrimination for visual stimuli, after Barlow.<sup>6</sup> Note the three regions of the data, with the intermediate regime corresponding to the Rose-de Vries law as described in the text.

the statistics of photon arrivals at the retina. The data could be interpreted by assuming that the observer “sees” the light whenever  $n$  photons arrive at his retina. Thus if the light has mean intensity  $I$  at the cornea, and the quantum efficiency from cornea to retina is  $\alpha$ , then the probability of seeing is

$$P(k > n) = \sum_{k=n}^{\infty} p(k|\alpha I), \tag{II.A.2.0}$$

where  $p(k|\alpha I)$  is the probability of  $k$  photons arriving given the mean number  $\alpha I$ . For light sources obeying Poisson statistics,

$$p(k|\alpha I) = e^{-\alpha I} \frac{(\alpha I)^k}{k!}. \tag{II.A.2.0}$$

The quantum efficiency  $\alpha$  is unknown, but the shapes of the curves  $P(k > n)$  vs.  $\log I$  depend *only* on  $n$ , the “threshold” photon number, as shown in Fig. II-2.

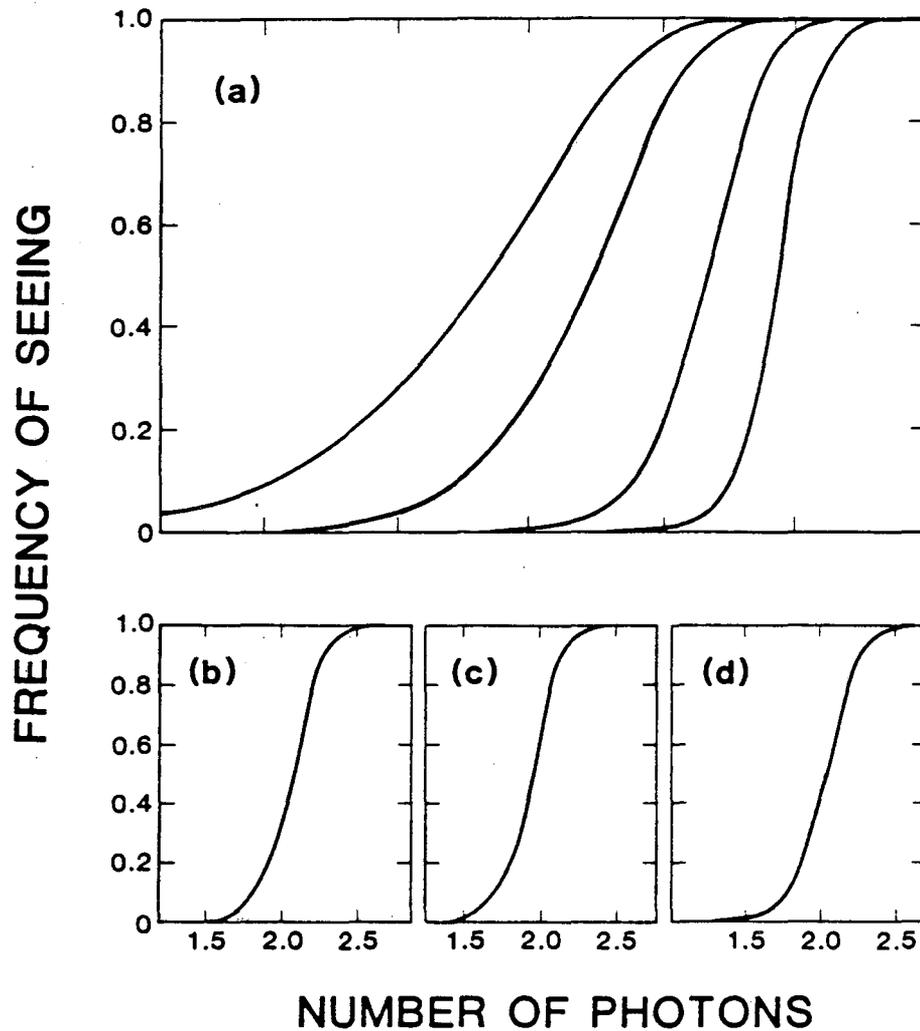


FIGURE II-2 Photon statistics and the frequency of seeing.<sup>7</sup> (a) The probability that more than  $n$  photons are absorbed by the retina given a flash of mean intensity  $I$  at the cornea. Note that curves for different values of  $n$  are very different, while changes in the quantum efficiency of the eye simply translate the curves from left to right. (b-d) Frequency of seeing curves for Hecht, Shlaer, and Pirrene, respectively, with fits to the curves of part (a).

The excellent fits of the data to these curves allow an estimate of  $n$ , which was found to range from 2 to 8 depending on the reliability of the observer's judgement.<sup>9</sup> Such "frequency of seeing curves" for Messrs. Hecht, Shlaer, and Pirenne are shown in Fig. II-2. Note that although Hecht requires more photons at his cornea because of his greater age; the apparent number of photons required at his retina is no different

from his younger colleagues.

Two historical points about the frequency of seeing experiment should be noted. First, M.H. Pirenne, who participated in these experiments, was L. Rosenfeld's student; we recall that Bohr and Rosenfeld were responsible for the first rigorous analysis of quantum limits to measurement of the electromagnetic field. Second, the idea of using Poisson fluctuations to characterize a biological system was applied in the same years to a very different problem.<sup>10</sup> Delbrück and Luria showed that the development of antibiotic resistance in bacteria resulted from spontaneous mutations, rather than adaptation, these two phenomena being distinguished by their effects on the statistics of a growing bacterial population. The fact that these spontaneous mutations occur on a reasonable time scale means that mutant bacteria can be selected in the same way as mutant strains of fruit flies; the Delbrück-Luria experiment thus marked the start of bacterial genetics, which formed the basis for the development of modern molecular biology.

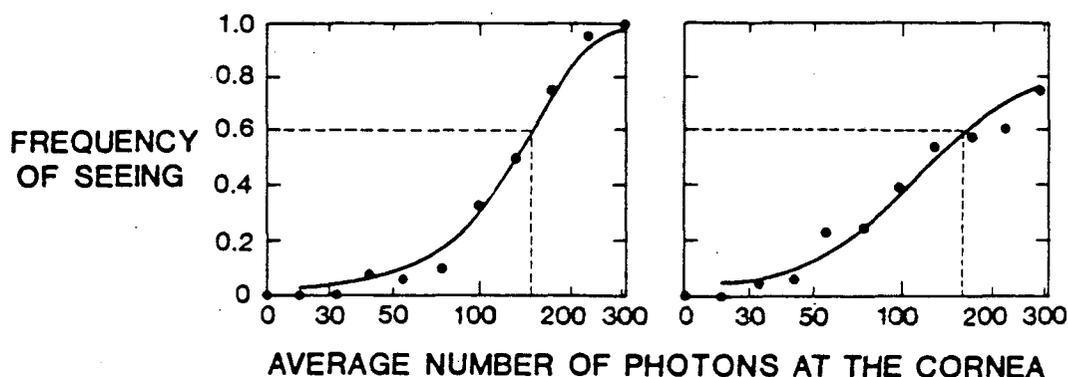


FIGURE II-3 Frequency of seeing curves, as in Fig. II-2, but comparing the results of (left) Poisson and (right) non-Poisson light sources in the same subject; adapted from Teich *et al.* (Refs. 11 and 12), with lines drawn to guide the eye.

Although the agreement between the experimental frequency of seeing curves and the theory of photon statistics is impressive, it is possible that many more photons actually reach the retina than the apparent number measured by these experiments,

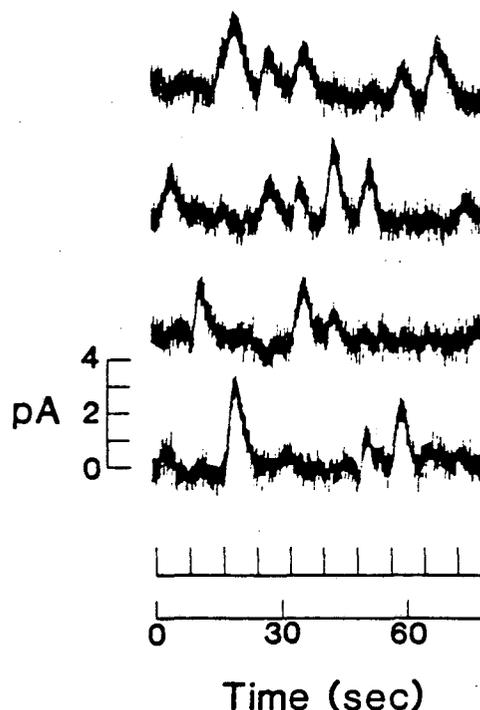
and hence that the visual threshold is limited by another source of fluctuation. This possibility was eliminated convincingly by the experiments of Teich *et al.*,<sup>11,12</sup> published in 1982, who repeated the frequency-of-seeing experiment with modulated laser light, a non-Poisson source. The results, summarized by the comparison of Poisson and non-Poisson curves in Fig. II-3, show that the light source with the greater variance in the photo-count distribution is less detectable for a given level of reliability, as we would expect if the statistics of photon arrivals at the retina were a major source of noise in the system. Further studies show that, for fixed mean intensity, the reliability of the visual observation declines as the variance of the photo-count distribution increases. Thus our ability to see a weak visual stimulus is indeed limited by the laws of quantum mechanics.

### *3. Single photon signals from individual receptor cells*

More recently, the psychophysical results have been confirmed qualitatively by recordings from individual photoreceptor cells.<sup>13</sup> As illustrated in Fig. II-4, the electrical current generated by these cells shows distinct "bumps" in the presence of dim light, and their occurrence follows the Poisson statistics expected if the each bump registers the arrival of an individual photon. By comparing the average single photon response with the spectrum of continuous current noise observed in the dark,<sup>14</sup> it is possible to estimate the reliability of the single photon response, and in this way the relation between psychophysical and physiological results can be made quantitative.

The calculational methods for making such reliability estimates are well known<sup>15,16</sup> but have not yet been applied to the photocurrent records from single receptor cells. In addition, many applications of these techniques to the sensory systems<sup>16</sup> suffer from a lack of clarity in dealing with continuous time-dependent signals.

FIGURE II-4 Single photon signals from individual receptor cells. This record shows the photocurrent generated by single rod outer segments in the retina of the toad, the stimulus being a sequence of dim flashes marked by ticks at the bottom of the figure.<sup>13</sup> Note that some flashes elicit no response, others a response  $\approx 1 \text{ pA}$ , and still other a response twice this large. The statistics of these null, unit, and double responses are consistent with the Poisson arrival of photons at the cell, the unit response corresponding to one photon. The absolute cross-section for producing the single-photon response is also in agreement with the pigment content of the cell, as discussed in Ref. 13.



More generally, the issue of reliability in the response of a single cell has not been discussed as quantitatively as one might like, particularly in relation to the idea that sensory systems possess a “threshold” for their response; this important issue is discussed in Appendix A with reference to auditory receptors, while in the following paragraphs I will present explicit calculations for the case of photoreceptors.

Imagine a record of the receptor cell current on the time interval  $0 < t < T$ . Our task is to decide from this record whether or not a photon was absorbed at time  $t = 0$ . In order to make this decision we must know the probability distributions for the time-dependent function  $J(t)$  which describes the current record. A continuous record, however, is described by an infinite number of variables (*e.g.*, the current at each instant of time) so that in integrating over the probability distribution we must integrate over an infinite number of variables, which amounts to evaluating a functional integral.<sup>17</sup> As suggested by Feynman<sup>17</sup> the functional integral, or path integral, thus provides a natural language for discussing problems in the theory of noise and

stochastic processes.

In the absence of photon absorption the current generated by the receptor cell exhibits<sup>14</sup> approximately Gaussian,<sup>18</sup> stationary fluctuations, and has a spectral density  $S(\omega)$ . This means that each Fourier component of the current record, denoted  $J(\omega)$  such that

$$J(t) = \frac{1}{\sqrt{T}} \sum_{\omega} e^{-i\omega t} J(\omega), \quad (\text{II.A.3.1})$$

is an independent Gaussian variable with a variance given by  $S(\omega)$ . Thus if there was no photon absorption event, the probability distribution for the function  $J$  is just

$$P_{-}[J] = K \exp \left\{ -\frac{1}{2} \sum_{\omega} \frac{|J(\omega)|^2}{S(\omega)} \right\}, \quad (\text{II.A.3.2})$$

where  $K$  is a normalization constant. Similarly, if there was an event, the current record is chosen from another probability distribution

$$P_{+}[J] = K \exp \left\{ -\frac{1}{2} \sum_{\omega} \frac{|J(\omega) - J_0(\omega)|^2}{S(\omega)} \right\}, \quad (\text{II.A.3.3})$$

where  $J_0(\omega)$  is the Fourier transform of the average single photon current response. A judgement of which probability distribution is more likely to have produced a given current record may be made by forming the "likelihood ratio"<sup>16</sup>

$$\lambda = \ln \left\{ \frac{P_{+}[J]}{P_{-}[J]} \right\}. \quad (\text{II.A.3.4})$$

If the occurrence of an event has greater probability given the current record than the absence of an event, then  $\lambda$  is positive.

In the presence of an event the probability distribution for the likelihood ratio is

$$p_+(\lambda) = \int [dJ] \delta \left\{ \lambda - \ln \left[ \frac{P_+[J]}{P_-[J]} \right] \right\} P_+[J], \quad (\text{II.A.3.5})$$

where  $\int [dJ]$  denotes functional integration, while if the photon is not present the likelihood ratio is distributed according to

$$p_-(\lambda) = \int [dJ] \delta \left\{ \lambda - \ln \left[ \frac{P_+[J]}{P_-[J]} \right] \right\} P_-[J]. \quad (\text{II.A.3.6})$$

These integrals are Gaussian, and may be done exactly,<sup>17</sup> to give

$$p_+(\lambda) = (M/\pi)^{1/2} e^{-M(\lambda-M)^2} \quad (\text{II.A.3.7a})$$

$$p_-(\lambda) = (M/\pi)^{1/2} e^{-M(\lambda+M)^2}, \quad (\text{II.A.3.7b})$$

where

$$M = \sum_{\omega} \frac{|J_0(\omega)|^2}{S(\omega)}, \quad (\text{II.A.3.8})$$

so that

$$\lim_{T \rightarrow \infty} M = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_0^{\infty} dt_1 \int_0^{\infty} dt_2 \frac{e^{j\omega(t_1-t_2)}}{S(\omega)} J_0(t_1) J_0(t_2). \quad (\text{II.A.3.9})$$

Assuming that the *a priori* probability of photon arrival is one half, the probability of making a correct judgement is

$$p_c = \frac{1}{2} \int_0^{\infty} d\lambda p_+(\lambda) + \frac{1}{2} \int_{-\infty}^0 d\lambda p_-(\lambda) \quad (\text{II.A.3.10})$$

$$= 0.5 + \frac{1}{2} \Phi(M^{3/2}), \quad (\text{II.A.3.11})$$

where  $\Phi(x)$  is the probability integral  $\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}$ .

Using the data cited above we can evaluate  $M$  through Eq. (II.A.3.9) and determine the probability that an ideal observer could correctly distinguish between a single photon and noise by recording the current from a single photoreceptor cell. The value of  $M$  obtained in this way is of order ten or larger, corresponding to  $p_c \geq 0.99$ . In this precise sense, the visual system is capable of counting single photons. Details of the calculation of  $M$ , as well as some further issues regarding the reliability of the single photon response, are discussed in Appendix B.

#### 4. Models for quantum counting

The time course of the current generated by vertebrate photoreceptor cells in response to the absorption of a single photon is quite slow and gradual, being described approximately by<sup>19</sup>

$$J_0(t) = A(kt)^n e^{-kt}. \quad (\text{II.A.4.1})$$

This time course is what we would obtain if the absorption of a photon by rhodopsin triggered a sequence of chemical reactions such that the last reaction consisted of opening or closing the ion channels in the cell membrane which control the flow of current into the cell. Such a "chemical cascade" is schematized in Fig. II-5; several issues regarding these models have recently been reviewed in one volume.<sup>20</sup> Similar models have been proposed for amplification in those biological systems which sense the concentrations of various chemicals in the cellular environment, such as hormones.<sup>21</sup> In each case, there is considerable experimental evidence to support the occurrence of a chemical cascade, although it may be argued that in no system have

the events of the cascade been shown to be the primary amplification steps in the sensory process. Thus there is still controversy in the case of photoreceptors as to whether the time course of the chemical events is consistent with that of the photocurrent, and whether the chemical events are involved in the amplification itself or only in its regulation (cf. Ref. 20).

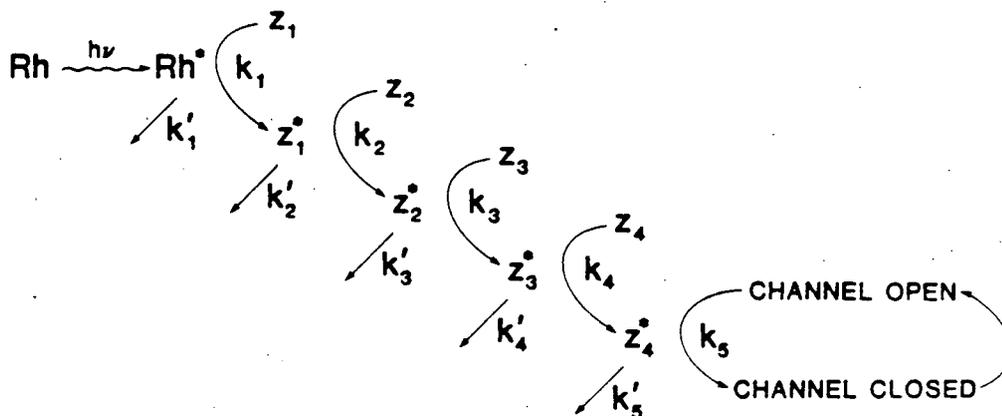


FIGURE II-5 A chemical amplifier in vision? A general version of the chemical cascade model, in which photon absorption by the visual pigment rhodopsin initiate a sequence of reactions, each of which catalyzes the next; the last step in the sequence consists of closing ion channels in the receptor cell membrane. The  $n^{\text{th}}$  molecule in the cascade lives for an average time  $(k'_n)^{-1}$  and catalyzes  $k_n$  reactions per second, so that the average gain  $\langle g_n \rangle = k_n/k'_n$ .

Any theory of amplification in vision must account for how the receptor is able to detect single photons with such remarkable reliability. It has been suggested that the chemical cascade can provide a natural explanation of this phenomenon, since a single photo-isomerized rhodopsin at the beginning of the cascade will result in many ( $\approx 10^4$ ) molecules being produced at the end of the cascade (cf. Ref. 20). This argument for large gain does not answer the question of reliable detection because it does not consider the statistical reliability of the gain itself.

Somewhat fancifully we may consider a large room full of a gas which undergoes some highly exothermic chain reaction, and we imagine that this reaction can be

initiated by photon absorption. Such a room may explode after the absorption of a single photon, but it is difficult to know precisely how many photons caused this explosion. This implies, in particular, that such a detector does not provide reliable discrimination among one or two photons; equivalently we may say that the response to each photon has a large variance. In fact, vertebrate photoreceptors do allow reliable discrimination among single and double photon arrivals, with the variance in the single photon response being of order ten percent or less.<sup>13</sup> Thus the exploding room, in spite of its sensitivity, does not constitute a good model for vision.

The reliability of the gain provided by the chemical cascade may be calculated as follows. At the  $i^{\text{th}}$  stage in the cascade, one molecule at the input is converted to  $g_i$  molecules at the output; in general the  $g_i$  are random variables whose statistics are determined by the detailed kinetics of the chemical events. The output of the entire  $n$ -stage cascade is therefore

$$G = \prod_{i=1}^n g_i \quad (\text{II.A.4.2})$$

output molecules per input photon. The relative variance in this gain is given by

$$\sigma^2 = \frac{\langle (\delta G)^2 \rangle}{\langle G \rangle^2} = \sum_{i=1}^n \frac{\langle (\delta g_i)^2 \rangle}{\langle g_i \rangle^2}, \quad (\text{II.A.4.3})$$

assuming that the fluctuations in gain are independent at each stage.

Consider first the case in which the gain at each stage varies according to the Poisson distribution. Thus we might imagine that each stage of gain is achieved by an enzyme which is activated for a period of time  $T$ , during which it catalyzes some reactions at a rate  $\lambda$ . The probability that  $g$  reactions occur for each activated enzyme is then

$$p(g) = e^{-\lambda T} \frac{(\lambda T)^g}{g!}, \quad (\text{II.A.4.4})$$

so that

$$\langle (\delta g)^2 \rangle = \langle g \rangle = \lambda T. \quad (\text{II.A.4.5})$$

If we look at the fluctuations in gain through the entire cascade we obtain from Eq. (II.A.4.3)

$$\sigma^2 = \sum_{i=1}^n \frac{1}{\langle g_i \rangle}. \quad (\text{II.A.4.6})$$

For fixed total gain  $G = \Pi \langle g_i \rangle$ , the minimum fluctuations are obtained when all the  $\langle g_i \rangle$  are equal; in this case

$$\sigma^2 = nG^{-1/n}. \quad (\text{II.A.4.7})$$

Two aspects of the Poisson model should be noted. First, the reliability of the gain decreases when it is spread over more stages; the optimum device therefore consists of a single stage of very high gain rather than many stages of intermediate gain. Second, the fluctuations in gain are much larger than might have been expected. Since we have assumed Poisson statistics, the expected result was that the relative variance in the gain is of order  $G^{-1}$ . Equation (II.A.4.7) shows that this obtains only in the case of a single stage. Because the fluctuations at each stage are not summed but *multiplied* by the cascade, the statistics of the whole cascade are very different from those of the individual stages. In particular, since the visual cascade is usually assumed to comprise four or more stages,<sup>19</sup> a gain of  $10^4$  results in a relative variance of greater than 0.4 and is thus inconsistent with the observations of relative variances of ten percent or less.<sup>13</sup>

Actually the statistics of a chemical cascade are much worse than indicated by the Poisson model. Consider again an enzyme which is activated at time  $t = 0$  and catalyzes a reaction at rate  $\lambda$ . The deactivation of the enzyme is itself a chemical reaction and is therefore statistical in nature. If the mean lifetime of the active enzyme is  $T$ , the probability that it will live for a time  $t$  is given by

$$p(t) = T^{-1}e^{-t/T}, \quad (\text{II.A.4.8})$$

assuming that the deactivation reaction occurs in a single step, as illustrated in Fig. II-5. The probability that the active enzyme will catalyze  $g$  reactions in its lifetime is therefore not given by Eq. (II.A.4.4) but rather by

$$\begin{aligned} p(g) &= \int dt p(t) p(g|t) = \int \frac{dt}{T} e^{-t/T} e^{-\lambda t} \frac{(\lambda t)^g}{g!} \\ &= \frac{1}{1+\lambda T} \left[ \frac{\lambda T}{1+\lambda T} \right]^g. \end{aligned} \quad (\text{II.A.4.9})$$

This exponential distribution of gain is contrasted with the Poisson distribution in Fig. II-6. The exponential distribution always has a variance equal to twice the square of the mean,

$$\langle (\delta g)^2 \rangle = 2 \langle g \rangle^2, \quad (\text{II.A.4.10})$$

so that unlike the Poisson distribution the exponential distribution does not become relatively narrow for large means.

A simple picture of chemical activation and deactivation in the cascade thus leads to exponential distributions for the gain at each stage, and from Eq. (II.A.4.3) this determines the relative variance of the cascaded gain to be

$$\sigma^2 = 2n, \quad (\text{II.A.4.11})$$

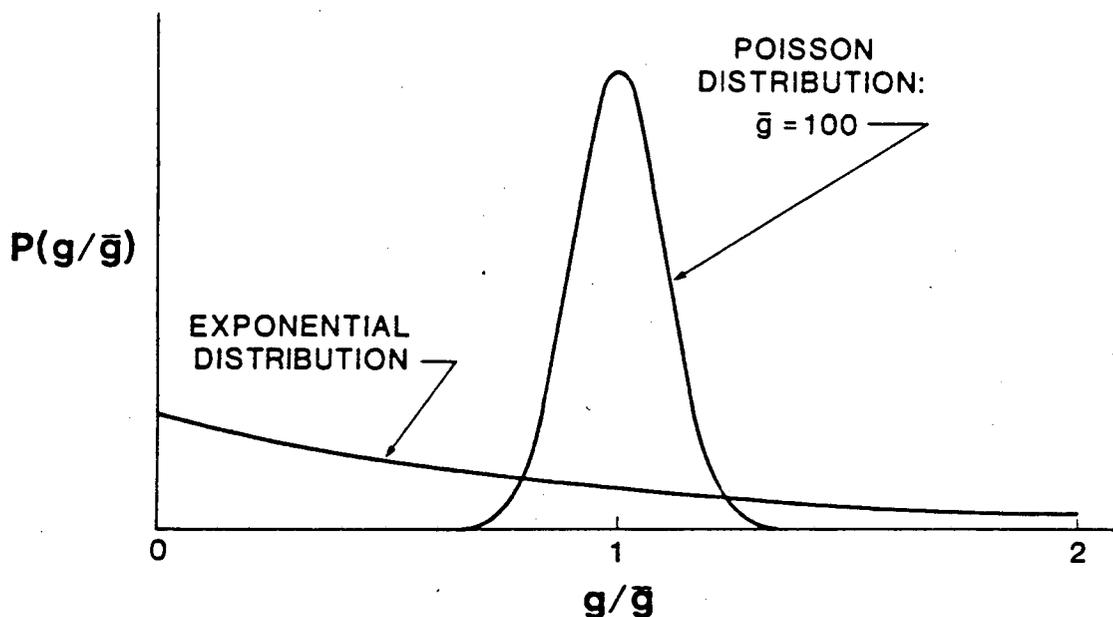


FIGURE II-6 Statistical aspects of a chemical amplifier. A comparison of Poisson and exponential distributions, with the same mean, illustrates why the gain of a chemical cascade is not reliable, even in the limit of large gain.

in clear disagreement with experiment. Thus the most straightforward implementation of the chemical cascade is completely unreliable, much like the exploding room example discussed previously, and cannot account for the quantum-limited performance observed in the visual system. This situation is analogous to that found in photomultipliers,<sup>22</sup> where the individual electron multiplication events are Poisson, but the inhomogeneities of the multiplier surfaces (here the distribution of enzyme lifetimes) can lead to highly non-Poisson, even exponential distributions for the photocurrent.

The example of the chemical cascade provides the first suggestion that conventional chemical mechanisms cannot describe quantum-limited behavior in biological systems. To determine whether or not this suggestion can be generalized--can *any* "chemical" model describe quantum phenomena in biological systems?--we must make more rigorous use of what has been learned in recent years about the quantum mechanics of macroscopic systems.

## B. Modern views of macroscopic quantum mechanics

### 1. Schrödinger's equation for a ton of brass

Consider a one ton brass bar. This system has many ( $\approx 10^{28}$ ) degrees of freedom. Of these, a few may be singled out as decidedly macroscopic, namely the lowest order modes of bending and sound propagation within the bar. Typically these modes are described by classical physics, which approximates them by harmonic oscillators with natural frequencies in the audible range. Nonetheless, there is no reason why we could not apply the principles of quantum mechanics to these normal modes, and quantize the classical oscillators if we wish.

For simplicity let us consider not the vibrational modes of a bar, but a mass  $M$  on a spring of stiffness  $\kappa$ . If we let the displacement of the mass from its equilibrium position be  $x$  and its momentum be  $p$ , then the Hamiltonian of the system is

$$H = \frac{1}{2M}p^2 + \frac{1}{2}\kappa x^2, \quad (\text{II.B.1.1})$$

and we can write the time-dependent Schrödinger equation for the wave function  $\psi(x)$ ,

$$i\hbar \frac{\partial \psi(x;t)}{\partial t} = -\frac{\hbar^2}{2M} \frac{\partial^2 \psi(x;t)}{\partial x^2} + \frac{1}{2}\kappa x^2 \psi(x;t). \quad (\text{II.B.1.2})$$

The important point is that the familiar quantum mechanics of a harmonic oscillator, usually applied to vibrations on the atomic and molecular scale, can be equally well applied to vibrations on the macroscopic scale. One way of understanding why this should work is to realize that the amplitude of a sound wave in a solid, for example, is a weighted sum of the displacements of the individual atoms in the solid. If we can apply quantum mechanics to the individual atoms we can apply it to the variables obtained simply by adding up the atomic variables.

More generally, the division into macroscopic and microscopic modes is arbitrary, and if it were not so cumbersome we could talk only about the motions of atoms, to which all agree quantum mechanics *must* be applied. If we did this, we would find that the equations of motion have solutions corresponding to sound waves, but of course since the whole approach is quantum mechanical these oscillations themselves would be quantized. Schrödinger's equation for the system will then have solutions corresponding to states of one, two, three, etc. quanta of sound waves, or phonons. The question we should be asking about the quantum mechanics of macroscopic systems is thus not "are quantum effects possible?," but rather "when are they important?." That is, when will the results of a calculation using a quantum oscillator differ significantly from those using a classical oscillator?

2. *When is the Heisenberg uncertainty principle important?*

The uncertainty principle states that if we make repeated measurements of the position and momentum of a particle, we will find finite variances  $\sigma_x^2$  and  $\sigma_p^2$  for these two quantities, and that the variances must obey the inequality

$$\sigma_x \sigma_p \geq \frac{1}{2} \hbar. \quad (\text{II.B.2.1})$$

As applied to an oscillator, this means that if we localize the particle to within  $\sigma_x$ , so that it has a potential energy  $\frac{1}{2} \kappa \sigma_x^2$ , it will have an uncertainty principle induced kinetic energy of  $\sigma_p^2 / 2M \geq \hbar^2 / 8M \sigma_x^2$ . Thus the energy of the system is

$$E = \frac{\hbar^2}{8M\sigma_x^2} + \frac{\kappa}{2} \sigma_x^2, \quad (\text{II.B.2.2})$$

The ground, or minimum energy state corresponds to

$$\sigma_x^2 = \langle (\delta x)_0^2 \rangle = \frac{\hbar}{2(M\kappa)^{1/2}}, \quad (\text{II.B.2.3a})$$

$$E_0 = \frac{1}{2}\hbar(\kappa/m)^{1/2}. \quad (\text{II.B.2.3b})$$

Thus even if we cool the oscillator to absolute zero, so that it is in its ground state, the uncertainty principle dictates that it undergo spontaneous fluctuations in position-- a sort of "quantum Brownian motion," termed zero-point motion.<sup>23</sup> In addition, the energy of the ground state is non-zero by amount proportional to  $\hbar$ , and this is the familiar zero-point energy  $\frac{1}{2}\hbar\omega$  with the resonant frequency given by  $\omega = (\kappa/m)^{1/2}$ .

The uncertainty principle is important whenever the amplitude of the zero-point motion is comparable to other motions in the system. We must apply quantum mechanics to molecules because (among other reasons) the atomic rearrangements which occur in chemical reactions may be only a few times larger than the zero-point motions of the atoms. Similarly we must apply quantum mechanics to, for example, a displacement sensor which receives signals comparable to its zero-point motions.

The case of the displacement sensor is particularly clear, and will be relevant to the sensory receptors of the inner ear to be discussed later. A "perfect" displacement sensor would provide a noiseless trace of position vs. time. By differentiating this trace with respect to time we can obtain the velocity of the particle, and hence its momentum. Again assuming the initial trace was noiseless, these two records would provide infinitely accurate, continuous readouts of position and momentum, with no variance in either variable. Of course this cannot happen, since it would violate the uncertainty principle. We conclude that a continuous readout of position vs. time must be noisy, and in fact this noise is just the zero-point motion found above.

In what sense is the quantum noise, corresponding in this case to zero-point motion, analogous to other physical noise sources, such as thermal noise (Brownian motion)? If we examine the probability distribution for the position of the oscillator, then in both cases--thermal and quantum--we find a Gaussian distribution. If we ask

for the variance of this Gaussian at finite temperature, where both thermal and quantum noise contribute, we obtain

$$\langle (\delta x)^2 \rangle = \langle (\delta x)_0^2 \rangle (2\bar{\nu} + 1), \quad (\text{II.B.2.4})$$

where the mean number of phonons

$$\bar{\nu} = (e^{-\hbar\omega/k_B T} - 1)^{-1}. \quad (\text{II.B.2.5})$$

This expression, the Planck formula, includes the zero-point motion at low temperatures and the classical Brownian motion at high temperatures, emphasizing that the two types of random motion should be viewed as aspects of the same phenomenon. Similar interpolation between classical (thermal) and quantum noise results in the generalization<sup>24</sup> of the Nyquist theorem for the spectral density  $S_V(\omega)$  of voltage fluctuations across a resistance  $R$ . Classically,

$$S_V(\omega) = \frac{k_B T}{\pi} R, \quad (\text{II.B.2.6})$$

while the full quantum result is

$$S_V(\omega) = \frac{\hbar\omega}{\pi} [\bar{\nu}(\omega) + 1/2] R. \quad (\text{II.B.2.7})$$

Recently it has been possible to directly test Eq. (II.B.2.7) by measuring the noise in superconducting devices.<sup>25,26</sup> The results demonstrate convincingly that the voltage noise does not fall to zero at zero temperature, as predicted by classical statistical mechanics, and hence that quantum noise is a *detectable* macroscopic manifestation of the uncertainty principle.

There is one important sense in which quantum noise is not like classical noise sources. The uncertainty principle only restricts measurements of *pairs* of

complementary variables, so that if we do not measure both elements of a complementary pair, there is no limitation; as an example we could imagine measuring only the momentum of a free particle. The idea of measuring only one element of the complementary pair is termed "quantum non-demolition," and such measurements in principle have no limits placed on them by non-relativistic quantum mechanics.<sup>27,28</sup> Thus the notion of quantum noise as equivalent to a classical noise source is not correct until we specify the nature of the measurement apparatus; related questions arise in the attempt to characterize quantum noise in amplifiers, as discussed in the following section.

Another example of how a displacement sensor must be noisy is provided by a simple free mass  $M$ .<sup>28</sup> If we localize this mass to a region  $\Delta x$  we impart to it a momentum uncertainty  $\hbar/2\Delta x$ , or a velocity uncertainty  $\Delta v = \hbar/2M\Delta x$ . After a time  $t$  this velocity uncertainty has added to the initial position uncertainty by an amount  $t\Delta v$ , so that the total uncertainty is

$$\Delta x(t) = \Delta x + \frac{\hbar t}{2M\Delta x}. \quad (\text{II.B.2.8})$$

If we minimize this uncertainty by varying  $\Delta x$  we obtain

$$\Delta x(t) \geq \sqrt{\frac{2\hbar t}{M}}, \quad (\text{II.B.2.9})$$

which is equivalent to a classical random walk with diffusion constant  $D_{eff} = \frac{\hbar}{M}$ .

To put the scale of "quantum diffusion" into biological perspective, we recall that many organisms<sup>29</sup> sense gravity and vibration with organs containing calcium carbonate crystals (*otoconia*) of dimension  $\approx 5\mu m^3$ . Since the specific gravity of these crystals is  $\approx 5$ , their mass is  $\approx 2.5 \times 10^{-14} kg$ , so that  $D_{eff} \approx 4 \times 10^{-21} m^2-s^{-1}$ . Over a

period of one second, the otoconia will therefore diffuse by about one angstrom even if thermal noise (Brownian motion) is neglected. In fact, the otoconial organs are sensitive to sub-angstrom stimuli, as are other biological systems (cf. section II.C.1). Thus it seems as though the uncertainty principle, as manifested through quantum noise, is significant in these systems, although we shall have to look more closely to decide if this argument can be made more rigorous.

The uncertainty principle applies not only to position and momentum, but to any pair of complementary variables, such as amplitude and phase. Again the argument is that if we make a precise measurement of the amplitude and phase of an oscillator, we can reconstruct its position and momentum, so that the uncertainty principle must restrict the accuracy of these measurements. If we measure amplitude in units  $N$  such that the average energy of the motion is  $N\hbar\omega$ , so that  $N$  is the number of phonons, then the amplitude-phase uncertainty relation is

$$\Delta N \cdot \Delta \phi \geq \frac{1}{2}. \quad (\text{II.B.2.10})$$

If we make continuous measurements of amplitude and phase, which are equivalent to the continuous readout of position discussed previously, then we will find the "standard quantum limits" analogous to the zero-point motion<sup>27</sup>:

$$\Delta N \geq \left(N + \frac{1}{4}\right)^{1/2}, \quad (\text{II.B.2.11a})$$

$$\Delta \phi \geq \frac{1}{2}N^{-1/2}, \quad N \gg 1. \quad (\text{II.B.2.11b})$$

Again it is important to put these relations in biological perspective, in this case by considering the electroreceptive ability exhibited by certain fish; many features of this unusual sensory system have been collected in recent reviews.<sup>30,31</sup> The "weakly electric" fish which live in the mountain lakes of South America and Africa generate

small electric fields which take the form of pulse trains or, in some species, quasi-sinusoidal waveforms. The distortions in field pattern which result from objects in the environment, as well as the fields generated by other fish, are detected by an array of specialized electroreceptors distributed over their body surface. The fish measures both the amplitude and phase of the field, since it can distinguish objects in the environment based on the relative reactive and resistive components in their electrical impedances, rather than just the magnitude of the impedance or a single quadrature phase.<sup>32</sup> The accuracy of this simultaneous amplitude-and-phase measurement is impressive: fish can detect amplitude changes<sup>33</sup> of  $\approx 10^{-7} V\text{-cm}^{-1}$  and phase shifts of  $\approx 10^{-4} \text{ rad}$ , the latter determined from the ability of the fish to phase-lock its oscillatory electric discharge to that of another fish,<sup>34</sup> a process which involves feedback from the receptors to the discharge organ.<sup>35</sup> Given the volume over which the receptors average and the bandwidths<sup>36,37</sup> through which they filter the signal, the threshold amplitude signal carries little more than  $k_B T$  of energy<sup>38,39</sup>; at a typical frequency of  $\approx 10^3 \text{ Hz}$  this corresponds to  $N \approx 6 \times 10^9$ . The standard quantum limit to phase accuracy is therefore  $\Delta\phi \approx 10^{-5} \text{ rad}$ , within an order of magnitude of the experimental result.

As in the case of the otoconial organs this argument is not rigorous, but the suggestion that biological sensors even approach the quantum limits to measurement is quite surprising. In the case of the displacement sensors of the inner ear, it is possible to present a rigorous argument; this will be done in Sections II.C below. In order to make such arguments, however, we must understand one more feature of macroscopic quantum mechanics, namely the role of amplifiers in the measurement process.

### *3. Bridging the gap between quantum and classical regimes: Amplifiers*

As long a signal is comparable to quantum noise, any observation of the detector will disturb it on account of the uncertainty principle. What we would like to do is

amplify the signal (and, inevitably, the quantum noise) to a level where it is above the intrinsic noise of the output device. Thus we might attach our detector to a meter whose needle moves in response to the signal, while in biological detectors the amplified signal is ultimately converted to a behavioral response such as the firing of a motor nerve.

All amplifiers make a transformation from "input modes" to "output modes." For example, we may have a harmonic oscillator which constitutes the detector and is the input mode, while the output mode may be the vibrations of the needle on the meter. To make these transformations precise, it is convenient to use the methods of second-quantization, so we recall some basic relations:

The case of a mass on a spring described by Eq. (II.B.1.1) will serve as prototype. The classical variables  $p$  and  $x$  are replaced by quantum operators and these operators obey the canonical commutation relation  $[x, p] = i\hbar$ . This relation implies the uncertainty principle through the general theorem

$$\Delta R_1 \cdot \Delta R_2 \geq \frac{1}{2} | \langle [R_1, R_2] \rangle |, \quad (\text{II.B.3.1})$$

where the operator variances are defined by

$$|\Delta R|^2 = \frac{1}{2} \langle RR^\dagger + R^\dagger R \rangle - \langle R \rangle \langle R^\dagger \rangle. \quad (\text{II.B.3.2})$$

The creation and annihilation operators for quanta of the oscillator are defined by

$$a^\dagger = (m\omega/2\hbar)^{1/2} (x - ip/m\omega), \text{ and} \quad (\text{II.B.3.3a})$$

$$a = (m\omega/2\hbar)^{1/2} (x + ip/m\omega) \quad (\text{II.B.3.3b})$$

respectively, and the number of quanta in the oscillator is represented by the operator

$$N = a^\dagger a. \quad (\text{II.B.3.4})$$

The commutation relations obeyed by the creation and annihilation operators are

$$[a, a^\dagger] = 1. \quad (\text{II.B.3.5})$$

This commutation relation is the precise mathematical correlate of the uncertainty principle, since through Eq. (II.B.3.2) we see that the operator variances are bounded by the commutators. These relations must therefore apply not only to the input mode of the amplifier but also to the output mode, or else we could violate the uncertainty principle by observing the position and momentum of, for example, the meter needle. This suggests that we focus upon the commutation relations of the output mode, and the condition imposed on the amplifier by the requirement that they be the same as those of the input mode. This line of argument was developed by Caves<sup>40</sup> who was able to settle the long-standing discussion about the quantum limits to the performance of linear amplifiers.<sup>41-43</sup> What follows is a presentation of the simplest version of his argument, and then an extension of the same type of argument to an example which demonstrates a strategy for quantum-limited measurement at finite temperature.

By a linear device we mean one in which the coordinate of the output mode is proportional to the coordinate of the input mode. In terms of the creation and annihilation operators this means that

$$b = M_{11}a + M_{12}a^\dagger + N \quad (\text{II.B.3.6a})$$

$$b^\dagger = M_{21}a + M_{22}a^\dagger + N^\dagger, \quad (\text{II.B.3.6b})$$

where  $b^\dagger$  and  $b$  create and annihilate quanta of the output mode, the matrix  $M$  allows for the most general linear transformation between input and output amplitudes, and the operators  $N^\dagger$  and  $N$  describe whatever noise is added to the signal by the amplifier. The fact that  $b$  and  $b^\dagger$ , like  $a$  and  $a^\dagger$ , are Hermitian conjugate to one

another requires

$$M_{11} = M_{22}^* \quad M_{12} = M_{21}^* \quad (\text{II.B.3.7})$$

These equations can be generalized to include multiple input and output modes.<sup>40</sup>

Equations (II.B.3.6), although they seem simple, actually provide a rigorous quantum mechanical description of amplifiers. In particular, if we impose the constraint  $[b, b^\dagger] = [a, a^\dagger] = 1$ , we obtain

$$[N, N^\dagger] = 1 - \det M. \quad (\text{II.B.3.8})$$

From Eq. (II.B.3.2), this means that the total noise variance added by the amplifier is

$$\frac{1}{2} \langle N^\dagger N + N N^\dagger \rangle \geq \frac{1}{2} |1 - \det M|. \quad (\text{II.B.3.9})$$

This result simplifies if we consider phase-preserving amplifiers, which have  $M_{12} = M_{21} = 0$  and cause phase shifts of the input to be mirrored as phase shifts of the output.<sup>44</sup> In this case  $\det M$  is simply the gain  $G$  of the amplifier, measured as the number of output quanta per input quantum, so that the added noise is

$$\frac{1}{2} \langle N^\dagger N + N N^\dagger \rangle \geq \frac{1}{2} |1 - G|. \quad (\text{II.B.3.10})$$

But this added noise is measured at the output; to refer it back to the input we simply divide by the gain, so that the added noise in terms of input quanta is given by

$$A \geq \frac{1}{2} |1 - G^{-1}|, \quad (\text{II.B.3.11})$$

or  $A \geq \frac{1}{2}$  for a high gain amplifier.

Thus a high gain phase-preserving amplifier must add an amount of noise which is equivalent to one-half quantum. But we have already seen that the incoming signal carries zero-point motion whose energy is one-half quantum [cf. Eq. (II.B.2.3)]: the

amplifier doubles the quantum noise in the signal. There are (at least) two things to be learned from this result. First, the quantum limit to amplitude and phase measurement is a factor of two worse than one might have predicted from zero-point motion alone. Second, if we are making measurements near the quantum noise limit then we must be using an amplifier whose noise performance is no worse than quantum mechanics demands it to be. In this sense we can define “perfect” amplifiers as those which add *only* the quantum minimum noise of Eq. (II.B.3.11), and conclude that measurement near the quantum limit requires a “perfect” amplifier.

The preceding analysis has concerned quantum-limited measurement in systems at absolute zero temperature, so that thermal noise may be neglected. Clearly this is not the case in biological systems, which are constrained to operate at or near 300K. Nonetheless, it is still possible to make quantum limited measurements. What follows is an analysis of one strategy for reducing the effects of thermal noise, together with a discussion of how quantum relations work their way into the problem and impose an absolute limit on the detection of small signals.

Imagine once again an oscillating mass on a spring, but this time being dragged through a viscous fluid so that it acquires a damping constant  $\gamma$ . The equation of motion becomes

$$m \frac{d^2 x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + \kappa x(t) = F(t) + \delta F(t), \quad (\text{II.B.3.12})$$

where  $F(t)$  is an external force applied to the system and  $\delta F(t)$  is the “Langevin force” which describes the effects of thermal noise. In the classical limit the Langevin force has a spectral density  $S_F(\Omega) = \gamma k_B T / \pi$ , as may be shown from the fluctuation-dissipation theorem.<sup>24</sup> Solving Eq. (II.B.3.12) for the statistics of  $\delta x$  in the presence of  $\delta F$  one finds the spectral density of displacement

$$S_x(\Omega) = \frac{\gamma k_B T / \pi}{(\kappa - m\Omega^2)^2 + (\gamma\Omega)^2}, \quad (\text{II.B.3.13})$$

where the root-mean-square fluctuations in any narrow band  $\Delta f$  are given by  $\delta x_{rms} = (4\pi S_x \Delta f)^{1/2}$  and the total fluctuations are

$$\langle (\delta x)^2 \rangle = \int \frac{d\Omega}{2\pi} S_x(\Omega) = k_B T / \kappa, \quad (\text{II.B.3.14})$$

as expected from the equipartition theorem.

A sinusoidal force of amplitude  $|F|$  applied at the resonant frequency  $\omega = (\kappa/m)^{1/2}$  produces a displacement amplitude of  $|x(\omega)| = |F|/\gamma\omega$ . If we make broad-band (or, equivalently, instantaneous) measurements of the displacement, then we see all of the displacement variance from Eq. (II.B.3.14), and the force will be reliably detected only if  $|F| \geq \gamma\omega\sqrt{k_B T/\kappa}$ . Since the spectrum of force noise is frequency independent, we can lower this threshold force by filtering.

One way to implement the filtering is by feedback, which also allows us to shift the resonance frequency of the oscillator. A schematic of how this might occur is shown in Fig. II-7. The idea is that an amplifier can be connected to the system, so that its input mode is the oscillation of the detector itself, and then the output (perhaps phase-shifted) can be applied back as a force on the detector. From a classical view we might imagine that this "feedback force" is in phase with velocity, and so constitutes a negative damping element of magnitude  $\eta$ . The equation of motion becomes

$$m \frac{d^2 x(t)}{dt^2} + (\gamma - \eta) \frac{dx(t)}{dt} + \kappa x(t) = F(t) + \delta F(t), \quad (\text{II.B.3.15})$$

and the same arguments as given above show that the threshold force becomes

$$F_{th}(\eta) = \gamma \omega \left( \frac{\gamma - \eta}{\gamma} \right)^{1/2} \sqrt{k_B T / \kappa}. \tag{II.B.3.16}$$

The effective reduction in noise is then

$$\frac{F_{th}(\eta)}{F_{th}(\eta = 0)} = \left( \frac{\gamma - \eta}{\gamma} \right)^{1/2}, \tag{II.B.3.17}$$

which is square root of the reduction in response bandwidth.

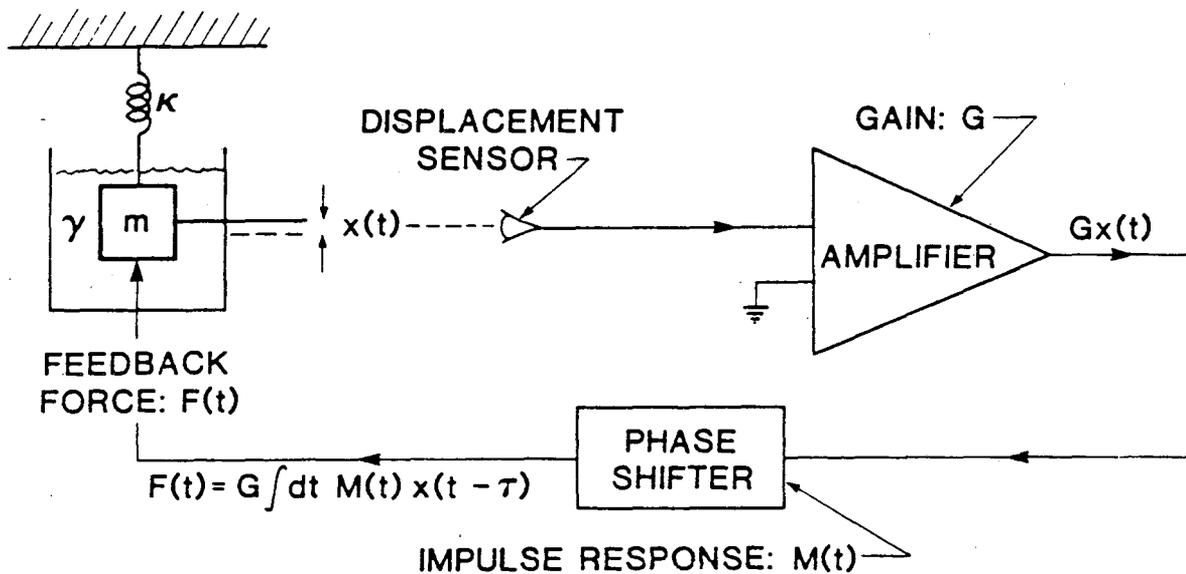


FIGURE II-7 Active feedback in a mass-spring system.

It would seem as if the reduction of noise by feedback has no limit, since we can narrow the bandwidth as far as we please, although we will pay for this in the time required for the detector to respond. We have, however, been considering an unrealistic model with no quantum noise and a noiseless amplifier. To give a consistent account of the quantum limits to noise reduction we must turn to a fully quantum treatment of the problem and keep in mind the following issues:

- [1] Active feedback can be used to synthesize reactive components (e.g. an effective mass) and thereby change the resonant frequency of a detector. To which frequency does the quantum noise (nominally  $\approx \hbar\omega$ ) correspond?
- [2] Active feedback can narrow the bandwidth of a detector and thereby reduce the effects of thermal noise, as outlined above; this may be described as lowering the effective temperature of the detector.<sup>45</sup> In the limit that the effective temperature is lowered below  $\hbar\omega/k_B$ , do we recover the appropriate quantum noise?
- [3] If the gain of the amplifier is increased, the system becomes unstable and oscillates. What are the quantum limits on the stability of these oscillations?

None of these issues, which are essential in understanding the role of amplifiers in finite-temperature measurements, could be resolved convincingly without using the techniques for describing amplifiers which were reviewed above. In addition, to give a rigorous account of these issues we must include a quantum treatment of the damping constant  $\gamma$  used in the classical analysis above. As will be described in more detail in Section III.A.2, this may be done by separating the many coordinates of the real system into the mode of interest and a "heat bath" to which the detector is coupled. Consider the model Hamiltonian

$$\mathbf{H} = \hbar\omega(a^\dagger a + 1/2) + i\Gamma(a^\dagger - a) + \mathbf{H}_{hb}, \quad (\text{II.B.3.18})$$

where  $\mathbf{H}_{hb}$  is the Hamiltonian of a heat bath and  $\Gamma$  is a coordinate of the bath which couples to the oscillator. Perturbation theory on the oscillator-bath coupling leads to the equations of motion<sup>46</sup>

$$\frac{da}{dt} = -i\omega a - \gamma a + \delta F \quad (\text{II.B.3.19a})$$

$$\frac{da^\dagger}{dt} = i\omega a^\dagger - \gamma a^\dagger + \delta F^\dagger, \quad (\text{II.B.3.19b})$$

where  $\delta F$  and  $\delta F^\dagger$  are quantum Langevin operators<sup>47</sup> with the properties

$$\int d\tau e^{i\Omega\tau} \langle \delta F^\dagger(t) \delta F(t-\tau) \rangle = 2\gamma\bar{\nu}(\Omega), \quad (\text{II.B.3.20a})$$

$$\int d\tau e^{i\Omega\tau} \langle \delta F(t) \delta F^\dagger(t-\tau) \rangle = 2\gamma(\bar{\nu}(\Omega) + 1), \text{ and} \quad (\text{II.B.3.20b})$$

$$\langle \delta F(t) \delta F(t') \rangle = \langle \delta F^\dagger(t) \delta F^\dagger(t') \rangle = 0, \quad (\text{II.B.3.20c})$$

where  $\bar{\nu}(\Omega) = (e^{\hbar\Omega/k_B T} - 1)^{-1}$ . If the output modes of the amplifier are created by  $b^\dagger$  and annihilated by  $b$ , then to simulate active feedback we must add to the Hamiltonian a term

$$\mathbf{H}_{\text{feedback}} = g(b^\dagger a + a^\dagger b), \quad (\text{II.B.3.21})$$

while the dynamics of the amplifier itself are described by

$$b(t) = \int dt' M(t-t') a(t') + N(t) \quad (\text{II.B.3.22a})$$

$$b^\dagger(t) = \int dt' M^\dagger(t-t') a^\dagger(t') + N^\dagger(t), \quad (\text{II.B.3.22b})$$

where  $N$  and  $N^\dagger$  are the operators which express the added amplifier noise. From these equations,

$$\begin{aligned} \frac{da(t)}{dt} &= -i\omega a(t) - \gamma a(t) - ig \int dt' M(t-t') a(t') \\ &\quad + \delta F(t) - igN(t), \end{aligned} \quad (\text{II.B.3.23a})$$

and

$$\frac{da^\dagger(t)}{dt} = i\omega a^\dagger(t) - \gamma a^\dagger(t) - ig \int dt' M^\dagger(t-t') a^\dagger(t')$$

$$+ \delta F^\dagger(t) - igN^\dagger(t). \quad (\text{II.B.3.23b})$$

We recall that the displacement  $q = [\hbar/2(m\kappa)^{1/2}]^{1/2}(a^\dagger + a)$ , while an external force  $F$  adds a term  $H_F = -Fq$  to the Hamiltonian. Then Eq's. (II.B.3.23) imply a response function

$$\frac{\bar{q}(\Omega)}{\bar{F}(\Omega)} = \frac{1}{2} (m\kappa)^{-1/2} \left\{ \frac{1}{\Omega + \omega + g\bar{M}(\Omega) - i\gamma} - \frac{1}{\Omega - \omega - g\bar{M}^\dagger(\Omega) - i\gamma} \right\}, \quad (\text{II.B.3.24})$$

where  $\bar{q}(\Omega) = \int dt e^{i\Omega t} q(t)$ , etc., while the Langevin terms in Eq's. (II.B.3.23) correspond to an effective noise force with spectral density

$$S_F^{eff}(\Omega) = 2\hbar(m\kappa)^{1/2} \int dt e^{i\Omega t} [\langle \delta F^\dagger(t) \delta F(0) \rangle + \langle \delta F(t) \delta F^\dagger(0) \rangle + g^2 \langle N^\dagger(t) N(0) \rangle + g^2 \langle N(t) N^\dagger(0) \rangle]. \quad (\text{II.B.3.25})$$

The force noise in the absence of feedback is given by

$$S_F^{(0)}(\Omega) = 2\hbar(m\kappa)^{1/2} \int dt e^{i\Omega t} [\langle \delta F^\dagger(t) \delta F(0) \rangle + \langle \delta F(t) \delta F^\dagger(0) \rangle] = 4\gamma(m\kappa)^{1/2} (2\bar{\nu}(\Omega) + 1), \quad (\text{II.B.3.26})$$

again in accord with the fluctuation-dissipation theorem.<sup>24</sup>

The correlation functions  $\langle N^\dagger(t) N(0) \rangle$  may be bounded by applying the quantum mechanical consistency condition that the commutation relations of the input and output modes be identical:

$$[b^\dagger(t), b(t')] = [a^\dagger(t), a(t')]. \quad (\text{II.B.3.27})$$

Applying this constraint to Eq's. (II.B.3.22) is simplified by passing to the Fourier representation; in this representation one obtains

$$[\tilde{N}^{\dagger}(\Omega), \tilde{N}(\Omega')] = [1 - \tilde{M}^{\dagger}(\Omega)\tilde{M}(\Omega')] [\tilde{a}^{\dagger}(\Omega), \tilde{a}(\Omega')]. \quad (\text{II.B.3.28})$$

Because all the noise processes are stationary, we must have  $\langle [\tilde{a}^{\dagger}(\Omega), \tilde{a}(\Omega')] \rangle = \tilde{C}(\Omega) 2\pi\delta(\Omega - \Omega')$ . The function  $\tilde{C}(\Omega)$  will be peaked near resonances of the response function in Eq. (7), and will be concentrated in a range of frequencies about these resonances comparable to their bandwidths.  $\tilde{C}(\Omega)$  must also be normalized to  $\int (d\Omega/2\pi) \tilde{C}(\Omega) = 1$  to preserve the equal time commutator  $[a(t), a^{\dagger}(t)] = 1$ . These relations, together with the generalized uncertainty principle of Eq. (II.B.3.2) imply that the spectral density of the amplifier noise  $N$  is

$$S_N(\Omega) \geq \frac{1}{2} |1 - \tilde{M}^{\dagger}(\Omega)\tilde{M}(\Omega)| |\tilde{C}(\Omega)|. \quad (\text{II.B.3.29})$$

From Eq. (II.B.3.25) we then obtain the effective spectral density of the force noise,

$$S_F^{eff}(\Omega) = S_F^{(0)}(\Omega) + (m\kappa)^{1/2} \hbar |\tilde{C}(\Omega)| g^2 |1 - \tilde{M}^{\dagger}(\Omega)\tilde{M}(\Omega)|. \quad (\text{II.B.3.30})$$

In the limit of a high-gain amplifier ( $|\tilde{M}| \gg 1$ ), we find that the added noise is

$$\Delta S_F^{eff}(\Omega) = (m\kappa)^{1/2} \hbar |\tilde{C}(\Omega)| |g\tilde{M}(\Omega)|^2. \quad (\text{II.B.3.31})$$

This is a particularly simple result because, from Eq. (II.B.3.24),  $g\tilde{M}$  is the "self-energy term" in the response function; that is,  $g\tilde{M}$  determines the change in the frequency response of the system when the feedback is applied. We thus arrive at the central conclusion: *the minimum noise added upon feedback from a high-gain amplifier is uniquely related to the change in frequency response achieved by the feedback.*

We can now answer the three questions outlined at the outset:

[1] If we want to shift the resonance frequency from  $\omega$  to  $\omega'$ , then we must have  $|g\tilde{M}| \approx |\omega - \omega'|$ . If the bandwidth of the resonance is narrow in the presence of feedback, then we may approximate

$$\int \frac{d\Omega}{2\pi} |\tilde{C}(\Omega)| |g\tilde{M}(\Omega)| \approx \int \frac{d\Omega}{2\pi} |\tilde{C}(\Omega)| (\omega - \omega')^2 = (\omega - \omega')^2, \quad (\text{II.B.3.32})$$

so that the frequency-integrated force noise,

$$\int \frac{d\Omega}{2\pi} \Delta S_F^{eff}(\Omega) \approx (m\kappa)^{1/2} \hbar (\omega - \omega')^2. \quad (\text{II.B.3.33})$$

Thus if feedback shifts the resonance down in frequency by a large amount, then the force noise added by the feedback amplifier becomes  $\approx \hbar (m\kappa)^{1/2} \omega^2 = \kappa \hbar (\kappa/m)^{1/2}$ , independent of the resonance frequency in the presence of feedback. Had we tried to detect the signal at  $\omega' \ll \omega$  in the absence of feedback, there would have been a quantum displacement noise (zero-point motion) of  $(\hbar/2\kappa)(\kappa/m)^{1/2}$ , while a force  $F$  at  $\omega'$  would produce a displacement  $F/\kappa$ , for an effective force noise of  $(\kappa\hbar/2)(\kappa/m)^{1/2}$ . Thus the amplifier noise “puts back” twice the quantum noise which was present before the feedback, and which was (nominally) removed by going to a much lower operating frequency.

[2] If we want to narrow the bandwidth of the detector from  $\gamma$  to  $\gamma - \eta$ , then we must have  $|g\tilde{M}| \approx \eta$ . The frequency-integrated force noise from the amplifier is therefore  $\approx (m\kappa)^{1/2} \hbar \eta^2$ . Again we can compare this to the situation in the absence of feedback, where the effective quantum force noise for a signal at resonance is  $(m\kappa)^{1/2} (\hbar\gamma^2/2)$ , so that in the narrow band ( $\eta \rightarrow \gamma$ ) limit the amplifier noise once again “puts back” twice the quantum noise which it filtered out.

[3] In the "bandwidth narrowing" configuration, instability results as soon as  $\eta \geq \gamma$ . Beyond this point the system is not a filter but an oscillator, emitting some stable signal of frequency  $\omega$  and bandwidth  $\Delta\omega$ . The effective spectral density of the force noise contributed by the amplifier which powers the oscillator is then  $S_F \geq (\hbar\gamma^2/\Delta\omega)(m\kappa)^{1/2}$ . If the oscillator is consuming a power  $P$ , then it is as if there were a force  $F_{eff} = (\gamma mP)^{1/2}$  across the dissipative element which determines the bandwidth  $\gamma$  in the absence of feedback. It is easy to see that  $\Delta\omega/\omega \approx (S_F\Delta\omega)^{1/2}/F_{eff}$ , so that the quantum limit to oscillator stability is given by  $\Delta\omega/\omega \geq \sqrt{\hbar\omega\gamma/P}$ . This result may be understood as follows: the energy  $E$  stored in the oscillator is dissipated in a mean time  $\gamma^{-1}$ , which means that  $E = P/\gamma$ ; thus the frequency stability is given by  $\Delta\omega/\omega \geq \sqrt{\hbar\omega/E}$ . But the frequency stability is just a measure of the phase noise,  $\Delta\omega/\omega \approx \Delta\phi$ , so that we obtain  $\Delta\phi \geq N^{-1/2}$ , where  $N$  is the number of quanta stored in the oscillator; this of course is precisely a factor of two more noise than the "standard quantum limit,"  $\Delta\phi \geq 1/2 N^{-1/2}$ . Once again the effect of the amplifier is simply to double the quantum noise, even when it serves to qualitatively change the dynamics of the system from small amplitude stability to instability.

#### 4. Conclusions: Quantum-limited measurement at finite temperature

The physical picture which emerges from this analysis is straightforward. Active feedback can be used to manipulate the frequency response of a system, and thereby improve the ratio of signal to thermal noise. Nominally we expect, from a Langevin approach, that these changes in frequency response could also reduce the effects of quantum noise, but if this were true we could use active feedback to circumvent the quantum limits to measurement. In fact, the self-consistent inclusion of the quantum limit on amplifier noise corrects this error, and yields the same quantum limits as would be obtained if the amplifier followed the detector with no feedback. Thus the limits to measurement imposed by the need to amplify a quantum signal up to the

classical level are the same whether the amplifier is used in series with the detector or as an essential part of the detector (in feedback), and this is as it must be.

Similarly, a Langevin approach suggests that, as the threshold of oscillation is approached and the bandwidth of the system narrows to zero, all the noise is filtered away and the frequency stability of the oscillation could be infinite. Again this conflicts with the uncertainty principle, which dictates a minimum phase noise, and again the self-consistent inclusion of the amplifier noise rectifies this error.

These conclusions emphasize that, although quantum noise may sometimes be effectively described by a Langevin "force operator," one must be careful in applying this approach to the quantum noise of amplifiers. This is because the quantum noise of amplifiers arises from a consistency condition between input and output and the apparent Langevin force changes whenever either input or output mode dynamics are changed. This may be contrasted with the Langevin forces describing thermal noise, which arise from coupling between the system and the heat bath, and therefore change only if this coupling is changed.

In practical terms, these results imply that amplifiers can be used in lieu of refrigerators--we can take many broad band detectors and apply feedback to synthesize narrow band detectors, thereby reducing the effects of thermal noise--but that, as in normal refrigerators, we can never freeze out the quantum fluctuations whose presence is dictated by the uncertainty principle.

In summary, the existence of quantum limits to measurement is one manifestation of quantum mechanics on a macroscopic scale. The key element in a quantum-limited device is the amplifier, and the use of amplifiers in feedback allows quantum limited measurements to be made even in the presence of large amounts of thermal noise. The validity of the uncertainty principle in such measurements is enforced by a

quantum limit on the performance of the amplifier itself, and only if this limiting performance is achieved can we reach the limits to measurement. The following sections discuss the evidence for the achievement of these limits in a biological detector system.

### C. Zero-point motion: Mechano-receptors of the inner ear

#### 1. Can we really hear sub-angstrom motions?

In the 1930's, von Békésy presented the first direct observations of motions in the inner ear, using stroboscopic microscopy of hearing organs (cochleae) dissected from cadavers.<sup>48</sup> In the same years, Autrum began the quantitative analysis of the mechanoreceptors on the legs of insects.<sup>49</sup> Both investigators obtained results suggesting that the threshold for a reliable response from the organism corresponds to displacements on the order of  $10^{-12}$  m. Time and again, this suggestion has been challenged by experimenters and theoreticians alike.<sup>50,51</sup> This section is an attempt to decide once and for all whether we can hear these sub-angstrom motions of the inner ear structures. The following sections will interpret the results--that the threshold for reliable response of the organism is below  $10^{-11}$  m--in terms of the quantum limits to measurement.

The cochlea of mammals is a complicated affair, being a spiral shaped cavity in the temporal bone. This cavity is divided into two fluid filled compartments by the *cochlear partition*, which consists of the *basilar membrane* and a set of structures which rest on top of that membrane. Among these structures are the *hair cells*, which are the receptor cells, and their *stereocilia*. It is the motion of the micron-sized stereocilia which elicits an electrical response of the hair cell,<sup>52</sup> which in turn modulates the firing of the auditory nerve fibers.

Sounds incident on the eardrum result in motions of the middle ear bones, which are connected to one of the fluid compartments in the cochlea. Because the cochlea is much shorter than the wavelength of sound ( $\lambda \approx 1\text{ m}$  at  $1\text{ kHz}$ , while the cochlea is  $\approx 35\text{ mm}$ ) in the fluid, inward motion of the window to one fluid compartment is balanced essentially instantaneously by outward motion of the window to other compartment. These opposing motions on either side of the partition result in a pressure difference across the basilar membrane, and the basilar membrane therefore moves; it is the pattern of this motion which von Békésy first observed.<sup>48</sup>

If we are willing to neglect the spiral curvature of the cochlea the physics of the situation is apparently not too complicated. We have two compartments sharing a common wall, the cochlear partition. All walls other than this common wall are assumed to be rigid, and since the wavelength of sound is long we may take the fluid itself to be incompressible. As long as the fluid motions are smaller than, for example,  $\approx 10\text{ }\mu\text{m}$  at  $1\text{ kHz}$ , the Navier-Stokes equations<sup>53</sup> for the fluid motion are effectively linear; if we can assign the basilar membrane a point impedance then the entire problem becomes linear. Allowing for smooth spatial variations of the membrane impedance, we obtain a wave-like solution traveling along the length of the cochlea, with a local wavelength which varies with the membrane impedance. In fact Békésy had shown by direct measurement that the stiffness of the membrane is greatest at the entrance to the cochlea and smallest at the opposite end.<sup>48</sup> From this fact alone one can show that the traveling wave envelope will peak near the cochlear entrance for high frequencies, and move further along the cochlea as the frequency decreases. Thus we expect a sorting of different frequencies to different cochlear locations, rather like an "acoustic prism".<sup>54</sup> Some details of this simple class of models for the macroscopic mechanics of the inner ear are reviewed in Appendix C.

When von Békésy looked through his stroboscopic microscope, he in fact observed such a traveling wave,<sup>48</sup> and the envelope of this wave exhibited the expected frequency sorting, which had first been predicted by Helmholtz.<sup>55</sup> Békésy himself used scale models to predict the pattern of motion from his stiffness data, and was satisfied by the agreement<sup>48</sup>; later investigators<sup>56-58</sup> showed that the same level of agreement could be obtained from model calculations. Although one could make sense out of all the mechanical data taken together, troubles began when one tried to compare the mechanical experiments with other measures of frequency sorting in the inner ear. In terms of the modern experiments on responses of the hair cells<sup>59</sup> and auditory nerve<sup>60</sup> it is clear that the frequency response at a point along the membrane is much broader in Békésy's observations than it is at the receptor itself; it would seem that we need a "second filter".<sup>61</sup>

The problem of the second filter, together with the tiny estimates of threshold displacements derived from the early experiments, led in the last two decades to application of modern experimental techniques to the measurement of basilar membrane displacement. Methods included Mössbauer probes,<sup>62-67</sup> capacitive probes,<sup>68,69</sup> and optical interferometry.<sup>70,71</sup> The great advance came in the 1980's when it was realized that the pattern of basilar membrane motion is, like the responses of the receptor cells, extremely sensitive to the physiological condition of the animal.<sup>67,71</sup> This result was presaged by differences between the pattern of motion in the same animal ante- and post-mortem.<sup>51,72,73</sup> Results on the pattern of motion under nearly ideal<sup>74</sup> conditions have now been reported in the cat<sup>71</sup> and the guinea pig.<sup>67</sup> Because far more data are available on the cat, I shall focus on these results in the attempt to determine the displacement at the threshold for a reliable response.

Khanna and Leonard<sup>71</sup> made interferometric measurements at the  $\approx 20$  kHz place in the cat. They directly observed  $10^{-10}$  m displacements of the basilar

membrane at a sound pressure of  $23 \text{ dB re } 2 \times 10^{-5} \text{ Nt--m}^{-2}$  ( $23 \text{ dB SPL}$ ) at the eardrum. It is important to note that the sound pressure is actually measured by a probe microphone placed very close to the eardrum, since the length of the ear canal is greater than the wavelength of sound at these frequencies, making extrapolations from sound pressure at the driver unreliable. The threshold for a reliable behavioral response of the animal at these frequencies is<sup>75-78</sup>  $-5 \text{ dB SPL}$  at the eardrum, so that if we assume linearity the threshold displacement is  $4 \times 10^{-12} \text{ m}$ . A number of observations indicate that the  $30 \text{ dB}$  linear dynamic range required for the validity of this estimate in fact exists for pure tones:

[1] Receptor cells in the inner ear produce a sinusoidal electrical response to sinusoidal sound pressure at the eardrum; the amplitude of this response is linear in the amplitude of the stimulus over the range of extrapolation.<sup>59,79</sup> The cells also produce a rectified DC potential in response to pure tones, and this rectified response grows as the square of the sound pressure.<sup>80</sup> This is consistent with linear growth of basilar membrane displacement driving a weakly non-linear transducer in the hair cell.

[2] At low frequencies the sinusoidal response of the hair cell voltage is reflected in a sinusoidal modulation of firing probability at the auditory nerve. The amplitude of this modulation also grows linearly for nearly  $30 \text{ dB}$  above behavioral threshold.<sup>81,82</sup>

[3] Linear growth of the sound pressure in the cochlear fluids near the place where Khanna and Leonard made their measurements.<sup>83,84</sup>

It is important to note, as Lynch *et al.*<sup>78</sup> emphasize, that the observation of combination tones and other two-tone interactions at these sound pressure levels is irrelevant to the question of linearity in the response to *single* tones. In fact the co-existence of single tone linearity and multi-tone non-linearity is one of the fundamental difficulties in understanding the non-linear behavior of the inner ear, and the evidence for pure

tone linearity should not be discarded in order to avoid this difficulty.

In summary, the evidence for linear responses of the mammalian cochlea at intensities between the behavioral threshold and the interferometric measurements is quite good, leading to an estimate of threshold displacement below  $10^{-11}$  m; the next generation of experiments should verify this extrapolation. It can also be checked against observations in the inner ear organs of other species.

Peake and Ling<sup>85</sup> have studied the vibration of the basilar membrane in the alligator lizard *Gerrhonotus multicarinatus* using the Mössbauer effect. To assess the effects of the surgery and probe placement on the condition of the inner ear, neural measurements were made with the Mössbauer source in place, both before and after the mechanical experiments.<sup>86</sup> The results of these experiments indicate that completely normal neural responses, both as regards frequency selectivity and threshold, can be obtained from the preparation used for the mechanical measurements.

In contrast to the mammalian cochlea, the basilar membrane of the alligator lizard does *not* exhibit a sorting of different frequencies to different places, although such organization is observed in neural and intracellular recordings.<sup>86</sup> Linearity of basilar membrane motion was found between 40 dB SPL and the highest levels used (no measurements were reported below 40 dB), while the thresholds reported for detectable changes in the firing rate of single afferent neurons range from  $\approx 10$  to 40 dB SPL.<sup>87</sup> By comparing these data one estimates the basilar membrane displacement at the "firing-rate threshold" to be  $(2.7 \pm 1.3) \times 10^{-11}$  m, where this estimate is based on the thresholds of 13 neurons; no systematic trends are seen as a function of frequency (or, equivalently, papillar place). In the case of the turtle--the only reptile for which behavioral data are available--Crawford and Fettiplace<sup>88</sup> have shown that the predominant neural response at behavioral threshold, for a range of frequencies that partially overlaps the sensitivity of the lizard, is a modulation of the neural firing

pattern and not a change in firing rate. Thus the behavioral threshold of the lizard is likely to be significantly lower (at least at some frequencies) than the firing-rate threshold, and the basilar membrane displacement at behavioral threshold is therefore probably below  $10^{-11} m$ .

Finally, we consider the responses of hair cells in the sacculus of frogs, which are sensitive to ground-borne vibrations.<sup>29</sup> In one species, clear changes in the firing pattern of the saccular neurons are observed when the entire frog moves by one-tenth angstrom.<sup>89</sup> Although there has been no quantitative analysis of the statistical properties of this neural response, so that it is impossible to define its reliability (cf. Appendix B), the general agreement with extrapolations from other inner ear organs is satisfying. We conclude that the vertebrate inner ear can detect displacements at or below  $10^{-11} m$ , as promised.

## 2. The ear at $T=0$ : Quantum noise and the threshold of hearing

The detector elements of the inner ear are the stereocilia, which are roughly cylindrical objects<sup>90,91</sup> of length  $L \approx 4 \mu m$  and radius  $R \approx 50 nm$ ; they consist of a crosslinked bundle of actin filaments.<sup>92</sup> From these facts we can estimate the mechanical properties of the cilium to be expected if the system is mechanically passive. All proteins,<sup>93</sup> including actin, have a density of  $\rho \approx 1.3 gm-cm^{-3}$  and a Young's modulus of  $Y \leq 2 \times 10^{10} Nt-m^{-2}$ . Thus a single stereocilium will have a mass  $m = \pi \rho R^2 L \approx 4 \times 10^{-17} kg$ . If the stereocilium is clamped at its base and free at its tip, then it will move in a cantilevered mode and have a stiffness  $\kappa = 3\pi YR^4/16L^3 \leq 10^{-3} Nt-m^{-1}$ . Indeed the cilia of many receptor cells are free-standing, and hinging of the base will only decrease the stiffness, so that this estimate is certainly correct as an upper bound.<sup>94</sup>

From these estimates we can give a lower bound to the zero-point motion of the stereocilium, namely

$$\langle (\delta x)^2 \rangle = \frac{\hbar}{2(m\kappa)^{1/2}} \geq (10^{-12} m)^2. \quad (\text{H.C.2.1})$$

Thus if we made broad-band amplitude and phase measurements of stereocilium motion in a fictitious ear capable of operating at absolute zero, the smallest signals we could detect would be displacements larger than  $10^{-12} m$ . In fact real ears at physiological temperatures do make amplitude and phase measurements, since the receptor cell produces a voltage which follows stereocilium displacement continuously,<sup>95</sup> and we can detect signals which are smaller than  $10^{-11} m$ . Thus there is not much difference between the lower bound on the quantum-limited signal and the upper bound on the threshold signal which is reliably detected. It is therefore certain that the inner ear makes measurements in a regime where quantum noise is important, and hence that hearing is a macroscopic quantum phenomenon.

The estimates presented in this section apply rigorously only to an ear in which thermal noise is somehow removed. In general we expect that the thermal noise is much larger than quantum noise in a room temperature device, so that an analysis of the quantum limits to hearing is not complete until we give an account of how this thermal noise is reduced and the quantum noise revealed.

### 3. The ear at finite temperature

The upper bound on the stiffness of a stereocilium determines the thermal noise displacements (Brownian motion) of the cilium which we would see if we made broad-band measurements of its motion. The result is that the mean-square displacement  $\langle (\delta x)^2 \rangle = k_B T / \kappa$ . Thus  $\delta x_{rms} = (k_B T / \kappa)^{1/2} \geq 2 \times 10^{-9} m$ , which is 40 dB above the threshold displacements reported for the inner ear.

To calculate the spectral density  $S_x$  of stereocilium Brownian motion, we need an estimate of its damping coefficient. An order of magnitude estimate may be obtained from the same hydrodynamic considerations which arise in the analysis of ciliary

beating.<sup>96</sup> For an object with the dimensions in the preceding section, we obtain  $\gamma \approx 10^{-10} \text{ Nt-s-m}^{-1}$ , assuming that the viscosity of the fluid surrounding the cilium is twice that of water.<sup>48</sup> From Eq. (II.B.3.13) above,

$$S_x(\omega) = \frac{\gamma k_B T / \pi}{(\kappa - m\omega^2)^2 + (\gamma\omega)^2}, \quad (\text{II.C.3.1})$$

and the root-mean-square fluctuations in a narrow bandwidth  $\Delta f$  are  $\delta x_{rms} = (4\pi S_x \Delta f)^{1/2}$ . From the parameter estimates above,  $\delta x_{rms} > 1.3 \times 10^{-12} \text{ m} (\Delta f / 1 \text{ Hz})^{1/2}$  for frequencies in the normal auditory range. If we are to detect  $10^{-11} \text{ m}$  displacements of single stereocilia, then the hair cell must possess a filter with a bandwidth of  $50 \text{ Hz}$  or less.

It might be supposed that much larger detection bandwidths could be tolerated by averaging over the many stereocilia on each hair cell, but this is not the case. At, for example,  $1 \text{ kHz}$ , fluid motion extends around the stereocilium through a boundary layer<sup>53</sup> of depth  $\approx 20 \mu\text{m}$  (for the viscosity of water), and objects within this layer--such as cilia on a single receptor cell, or even nearby cells--will be coupled through the viscosity of the fluid. Systems which are viscously coupled exhibit correlated Brownian motions,<sup>24</sup> and hence there will be little or no noise reduction produced by such spatial averaging. The fact that there is no significant noise reduction with spatial averaging predicts that the central nervous system needs to look only at a single auditory nerve fiber in order to reliably detect the threshold signal, and as discussed in Appendix A this is indeed the case.

If we cannot reduce the noise level by averaging perhaps we can raise the signal level by some means. Thus although the basilar membrane moves by only  $10^{-11} \text{ m}$ , the fluids in the immediate neighborhood of the stereocilia might move much more, possibly enough to overcome the thermal noise problem. The signal must create shear

in the fluid, however, and this takes power. In particular, maximum stimulation of the stereocilia will occur if the fluid near its tip moves at velocity  $\omega x_{fl}$  and the fluid near the base is at rest. If we wish to stimulate one hair cell, the best we can do is to distribute this shear force over one boundary layer depth  $\delta$ . From this picture, the power dissipation will be

$$P_D = \frac{1}{2} \pi \delta^2 \frac{\eta}{L} \omega^2 x_{fl}^2, \quad (\text{III.C.3.2})$$

where  $\eta$  is the viscosity. The power which enters the cochlea at threshold can be calculated from the sound pressures in the cochlear fluids<sup>83,84</sup> and the acoustic impedance<sup>78</sup> of the organ; the result<sup>38</sup> is that  $P_D \leq 6 \times 10^{-19} W$ . With the parameters given above, this determines the fluid displacement in the neighborhood of the cilia to be  $x_{fl} \leq 10^{-10} m$  at 1 kHz. This fluid displacement will result in a stereocilium displacement  $x_s$  given by

$$|x_s|^2 = \frac{(\gamma\omega)^2 x_{fl}^2}{(\kappa - m\omega^2)^2 + (\gamma\omega)^2}, \quad (\text{III.C.3.3})$$

for a broad-band signal-to-noise ratio of

$$SNR = \frac{\kappa x_{fl}^2}{k_B T} \frac{(\gamma\omega)^2}{(\kappa - m\omega^2)^2 + (\gamma\omega)^2}. \quad (\text{III.C.3.4})$$

If the cilium were naturally resonant at  $\omega$  this signal-to-noise ratio would be maximized. In fact, from the estimates of  $m$  given above, which depend only on geometry and protein density, both of which are well known, values of  $\kappa$  sufficiently small to generate a resonance at 1 kHz would result in Brownian motions of several microns. We must therefore reject a passive resonance model of the cilia by themselves. If we mass-load the cilia by attaching their tips to some other structure, we can

lower the effective thermal noise in the cantilevered bending mode but not in the other modes of motion, as illustrated in Fig. II-8. Since the receptor cell appears to measure angular displacement of the cilium at its base<sup>95</sup> these noise sources are equivalent, and they are also of comparable magnitude.<sup>97</sup> From these considerations, we may take  $m\omega^2 \ll \kappa$  hereafter, so that

$$SNR = \frac{\kappa x_{fl}^2}{k_B T} \frac{(\gamma\omega)^2}{\kappa^2 + (\gamma\omega)^2}$$

$$\leq \frac{\gamma\omega x_{fl}^2}{2k_B T} \approx -50 \text{ dB}, \tag{III.C.3.5}$$

where the upper bound in the last equation is obtained by optimizing  $\kappa$ . Thus, with fluid motion properly accounted for, the thermal noise problem actually becomes worse than first calculated.

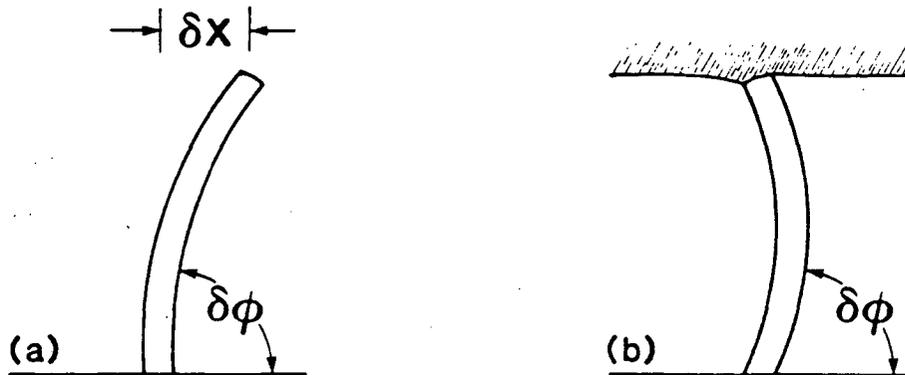


FIGURE II-8 Thermal noise and modes of stereocilium motion. The cantilevered mode (a) is that considered in the text. If the stereocilium were rigidly attached to the tectorial membrane which overlies it, the Brownian motion in the cantilevered mode would be reduced dramatically, but there would still be random motion in the flexing modes, such as that shown in (b). The displacements in different modes can be compared by examining the equivalent bending angle at the base of the stereocilium, as indicated.

The reason for the greater seriousness of the problem in this calculation is that the cilium is severely impedance mismatched to its fluid environment. The first calculation assumed that the cilium moved by  $10^{-11} m$ , which is comparable to the fluid displacement and therefore requires good impedance matching. But the high stiffness assumed in that calculation is inconsistent with such a good match; to improve the match we must make the cilium more compliant, so that it is resistance rather than stiffness dominated. We pay for this in increased Brownian motion, and there is an optimum which is calculated in Eq. (II.C.3.5).

Consideration of the matching problem is of interest for two reasons. First it provides a better estimate of the magnitude of the thermal noise level, and hence of the difficulties which the ear faces. Second it shows that, for example, rigidly linking many stereocilia together to form a stiffer object cannot solve the noise problem--although the broad-band Brownian noise will be lowered, the poorer impedance match will more than compensate any gain in signal-to-noise.

An alternative approach to overcoming stereocilium Brownian motion is by amplifying the motion of the basilar membrane. This will be disastrous if the Brownian motion of the membrane itself is significant. To estimate the Brownian motion of the membrane, we write the energy of the system in terms of the position and velocity of the membrane at each point in the cochlea, and then apply the Boltzmann distribution.

In conventional models of membrane mechanics (cf. Appendix C), the energy consists of three components: the kinetic energy of the basilar membrane, the kinetic energy of the cochlear fluids, and the potential energy of the membrane. Considering the kinetic energy terms, the basilar membrane velocity  $v(x)$  at the point  $x$  along the cochlea has the spatial Fourier representation

$$v(x) = \int \frac{dk}{2\pi} e^{ikx} V(k), \quad (\text{II.C.3.6})$$

so that the kinetic energy of the membrane

$$\begin{aligned} K.E. (\text{membrane}) &= (mb/2) \int dx v^2(x) \\ &= (mb/2) \int \frac{dk}{2\pi} |V(k)|^2, \end{aligned} \quad (\text{II.C.3.7})$$

where  $m$  and  $b \approx 100 \mu m$  are the mass per unit area and width of the membrane, respectively. The kinetic energy of the fluid can be written

$$K.E. (\text{fluid}) = (b/2) \int \frac{dk}{2\pi} \rho Q^{-1}(k) |V(k)|^2, \quad (\text{II.C.3.8})$$

where  $\rho \approx 1 \text{ gm-cm}^{-3}$  is the fluid density and  $Q(k)$  depends on the assumed geometry of the model (cf. Appendix C). Thus, if the fluid motion is primarily one-dimensional,  $Q(k) = hk^2$ , where  $h \approx 0.1 \text{ cm}$  is the effective height of the cochlear chambers, while if the motion is two-dimensional  $Q(k) = k \tanh(kh)$ . The total kinetic energy is therefore

$$K.E. (\text{total}) = (b/2) \int \frac{dk}{2\pi} [m + \rho Q^{-1}(k)] |V(k)|^2. \quad (\text{II.C.3.9})$$

According to the Boltzmann distribution, the probability of any particular configuration of the system is  $e^{-E/k_B T}$ , where  $E$  is the energy of the configuration. Applying Eq. (II.C.3.9) for the energy, the  $V(k)$  are Gaussian random variables; the variances are given by

$$\langle V(k) V(k') \rangle = \frac{k_B T \delta(k+k')}{b[m + \rho Q^{-1}(k)]}, \quad (\text{II.C.3.10})$$

so that

$$\langle v^2(x) \rangle = (k_B T / mb) \int \frac{dk}{2\pi} \frac{Q(k)}{Q(k) + \rho/m}. \quad (\text{II.C.3.11})$$

For  $Q(k)$  given above the integral may be evaluated as

$$\langle v^2(x) \rangle = (k_B T / 2bh^2\rho)(\rho h/m)^{3/2} \quad \rho h \ll m, \quad (\text{II.C.3.12a})$$

$$= (k_B T / bh^2\rho)(\rho h/m)^2 \quad \rho h \gg m. \quad (\text{II.C.3.12b})$$

Current models<sup>98</sup> often assume  $m \approx \rho h$ , so that the thermal fluctuations in basilar membrane velocity will be

$$\delta v_{rms} \approx (k_B T / bh^2\rho)^{1/2} \approx 2 \times 10^{-5} \text{ cm-sec}^{-1}. \quad (\text{II.C.3.13})$$

At a frequency of 1 kHz, these velocity fluctuations are equivalent to displacement fluctuations of  $\approx 10^{-10}$  m, or more than 20 dB above threshold. It may be shown that the correlation length for these fluctuations is  $\approx h$ , so that no reasonable spatial averaging could reduce their effect.

The same reasoning may be applied to the potential energy, which is

$$P.E. (\text{membrane}) = \frac{b}{2} \int dx C(x) z^2(x), \quad (\text{II.C.3.14})$$

where  $z(x)$  and  $C(x)$  are the displacement and stiffness per unit area of the membrane, respectively. In analogy to Eq. (II.C.3.10),

$$\langle z(x)z(x') \rangle = k_B TC^{-1}(x)\delta(x-x'). \quad (\text{II.C.3.15})$$

If we average over a region of length  $d$ , we will see displacement fluctuations of  $\delta x_{rms} = [k_B T / bdC(x)]^{1/2}$ , and  $C$  may be determined from the resonance condition  $\omega_o^2 m = C$ . Thus, at the position in the cochlea corresponding to  $\omega_o = 2\pi(1 \text{ kHz})$ , and with  $m \approx \rho h$  from above, a hair cell of diameter  $d \approx 10 \mu\text{m}$  will "see" a basilar

membrane Brownian motion  $\delta x_{rms} \approx 3 \times 10^{-10} m$ , or about 30 dB above threshold.

Whether we measure displacement or velocity, the thermal noise at the basilar membrane is large--too large to allow significant amplification of the membrane motion without making a serious problem much worse. The need to reduce the effects of this noise argues strongly for a filtering process subsequent to basilar membrane mechanics; essentially a "second filter" in the sense of Evans and Wilson.<sup>61</sup>

Even assuming perfect coupling through the cochlea, so that all the power entering the cochlea is available to drive fluid motion near the stereocilia, we have found that the thermal noise of the stereocilia is at least 40 dB above the threshold signal. This problem cannot be resolved by loading the cilium with supporting structures, by spatial averaging, or by broadband amplification of the signal. The only possible mechanism of noise reduction is filtering, and this filter must reside within the hair cell itself rather than in auxiliary structures such as the basilar membrane. The bandwidth of the filter calculated from the spectral density of displacement noise is  $\Delta f \approx 50 Hz$ , and this agrees with the threshold power since  $6 \times 10^{-19} W \approx 4k_B T \Delta f$ . The most careful measurements of the receptor bandwidth also agree with this prediction.<sup>88</sup>

The filtering required for the rejection of thermal noise cannot be the result of a passive resonance of the stereocilia for the reasons discussed above. We could, in principle, perform the filtering electrically after the mechanical signal has been transduced by the receptor cell, but this runs afoul of the transducer noise to be discussed in the following section. Thus the filtering must be done at the level of stereocilium mechanics, but not as the result of a passive resonance: we must have an active mechanical filter in the stereocilia.

At least one biological system is known to include an active mechanical filter, and this is the asynchronous insect flight muscle. In insects which have this type of flight muscle, the beating of the wing is not governed by the rhythm of neural impulses, but rather by an oscillation of the muscle itself.<sup>99</sup> If we extract the muscle and "relax" it by manipulating its ionic environment, the muscle is very stiff, in contrast to the more familiar vertebrate striated muscle. This stiffness results from filaments which run the whole length of the muscle, and their mechanical properties are thus in parallel with whatever active component is generated when the muscle is activated.<sup>100</sup> When the muscle is activated, the active component of the mechanical response has an amplitude which is roughly independent of frequency, but whose phase changes from "stiffness-like" to "mass-like" as frequency is increased; at some point the mass-like term generates a resonance with the passive stiffness, and this determines the characteristic frequency of the muscle.<sup>101,102</sup> These features of muscle mechanics, which are exactly those expected from an active filter, are summarized in Figs. II-9 and II-10.

The flight muscle provides a biological precedent for the mechanism which we require to understand the reduction of thermal noise in the inner ear. In fact it may also provide a fruitful analogy, since the major protein components of muscle are found in the stereocilia and the adjacent structures of the receptor cell.<sup>92,103-105</sup> Thus we imagine each stereocilium to be a small piece of muscle, filtering the mechanical signals applied to it by way of its active mechanical response. The system acts to reduce the effects of thermal noise in exactly the way described in Section II.B.3., making use of feedback to both narrow the bandwidth and shift the resonance frequency.

The important point is that an active mechanical filter, whatever its internal mechanism, is required if we are to make sense of the threshold signals which the inner ear can detect. The active filter hypothesis does not give, however, a complete

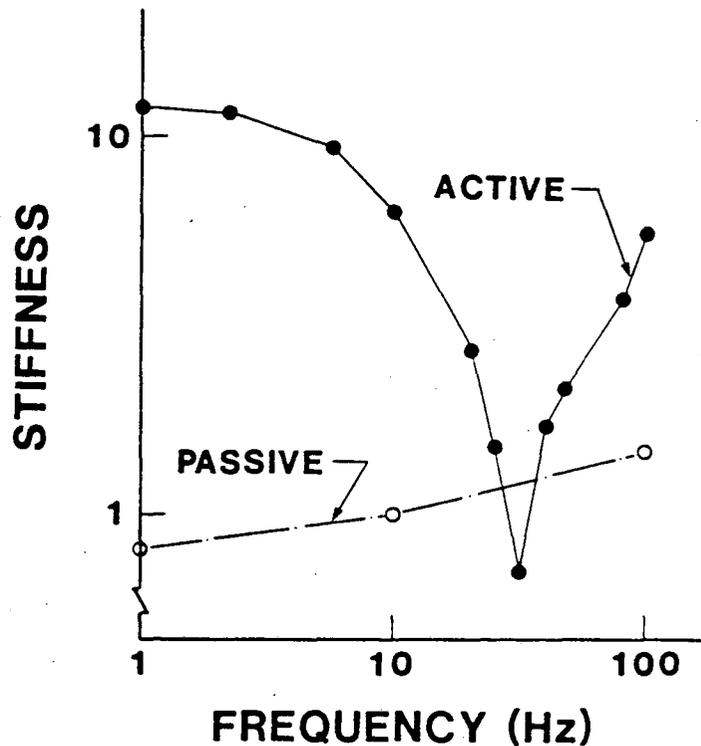


FIGURE II-9 Active mechanical filtering in a biological system--Insect flight muscle. Resonant behavior of the muscle in the active state is contrasted with the approximately elastic behavior in the relaxed (passive) state; data from the work of Wilson and White.<sup>101,102</sup>

account of the problem. First, we have not explained how the signal gets to the stereocilia from the basilar membrane. Clearly these two mechanical systems must be tightly coupled, or else much of the signal will be lost; recall that all the calculations of thermal noise have been based on the assumption that no such loss occurs. If tight coupling occurs, however, we will "see" the active mechanical resonance of the stereocilia when we look at the basilar membrane. In this way, as discussed in Appendix C, active filtering at the stereocilia will lead to a pattern of basilar membrane motion very different from that expected in a passive cochlea, and more in accord with modern experiments.

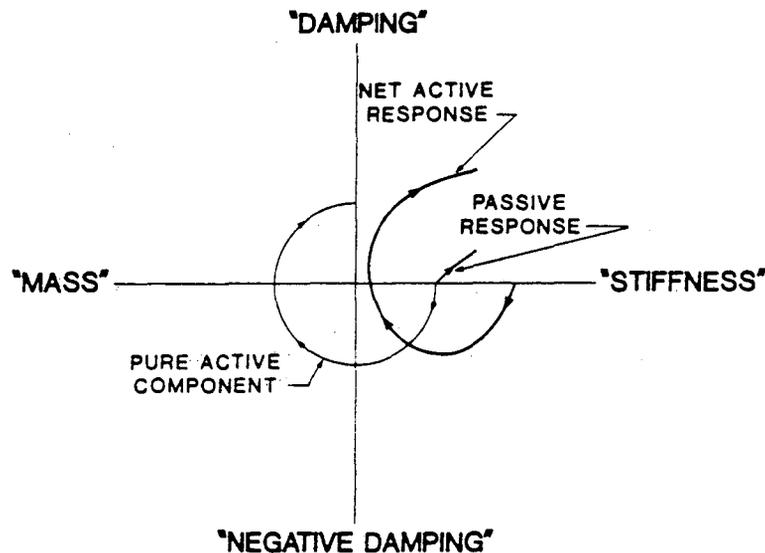


FIGURE II-10 Flight muscle filters--A schematic decomposition. Following Wilson, White, and Thorson,<sup>100-102</sup> we consider the Nyquist plot of the muscle response, that is a graph of the real part of the muscle "stiffness"  $F(\omega)/x(\omega)$  against the imaginary part for each frequency  $\omega$ ; frequency increases along the curves, in the direction of the arrows. The net active response is the result of a parallel combination of the true active response and the passive mechanics of the muscle filaments. Measurement of the latter in relaxed muscles allows unambiguous separation of the two components, as shown. Note that the active element has an amplitude which is roughly frequency-independent, but a phase which shifts from "stiffness-like" to "mass-like" and ultimately to "damping-like" as the frequency is increased. It is this unusual phase-frequency behavior in the active element of the muscle which produces the active resonance of Fig. II-9.

The second problem with the simple description of active filtering brings us back to quantum noise. If active filtering occurs in the inner ear, then since the signal being detected is comparable to the broad-band quantum noise of the passive system (cf. Section II.C.2) we know that the amplifier used in the feedback process must be quantum mechanically perfect (cf. section II.B.3). Thus the ear not only makes use of a novel mechanism to reduce its thermal noise, it possesses a device as noiseless as the uncertainty principle allows, and in this way it makes quantum-limited measurements in spite of thermal noise. It is the presence of this device, a quantum limited amplifier, which makes the perception of sound a macroscopic quantum phenomenon.

#### 4. Models for a linear amplifier

Many models for amplification processes in the sensory systems have been based on analogies to the behavior of nerve membrane. In this section I review these models and consider their noise performance in relation to the observation that the inner ear makes use of a quantum-limited amplifier. We shall see that the extant models cannot account for quantum-limited performance, and that this is a fundamental failure of any model which may be described in the conventional language of chemical kinetics. Examination of how the "chemical amplifier" fails defines the conditions on the internal mechanism of the amplifier which must be met if the amplifier is to exhibit quantum-limited performance.

The response of a patch of nerve membrane to small voltages is linear, and this linear response exhibits a resonance.<sup>106</sup> As the standing voltage across the membrane is stepped toward the threshold for initiation of an action potential, this resonance becomes sharper and sharper, approaching infinite sharpness and instability at the threshold itself; beyond threshold there is no stable small-signal response of the membrane. This resonant behavior is evidence of an apparent inductive component to the membrane impedance which, together with the intrinsic capacitance of the membrane, generates the observed resonance.

The apparent inductance of a biological membrane is believed to result from the presence of voltage-gated channels.<sup>107</sup> The simplest example of such a channel is a protein inserted in the cell membrane which can exist in two states, "open" and "closed." In the open state the channel allows ions to cross the membrane and contributes a conductance  $g$ ; in the closed state the channel allows no ions to pass and does not contribute to the membrane admittance. To have a voltage-gated channel, we assume that the rates for opening  $k_+$  and closing  $k_-$  of the channel may depend on the trans-membrane voltage  $V$ . Thus if  $n$  denotes the number of open channels in

the membrane and  $N$  denotes the total number of channels,

$$\frac{dn}{dt} = k_+(V)[N-n] + k_-n. \quad (\text{II.C.4.1})$$

The current through the channels is given by

$$I = gn. \quad (\text{II.C.4.2})$$

If we bias the membrane at a voltage  $V_o$  and consider the response of the channels to small changes  $\delta V$ , we obtain

$$n = n_o + \delta n, \quad (\text{II.C.4.3})$$

where

$$n_o = k_+(V_o)/[k_+(V_o) + k_-(V_o)], \quad (\text{II.C.4.4})$$

and

$$\frac{d(\delta n)}{dt} = -(1/\tau)[\delta n - (\frac{\partial n}{\partial V})\delta V], \quad (\text{II.C.4.5})$$

where

$$1/\tau = k_+(V_o) + k_-(V_o) \quad (\text{II.C.4.6})$$

$$\frac{\partial n}{\partial V} = \tau \left[ \frac{\partial k_+}{\partial V} (1 - k_+\tau) - \frac{\partial k_-}{\partial V} k_+\tau \right] \quad (\text{II.C.4.7})$$

Fourier transforming Eq. (II.C.4.5) allows us to find the apparent impedance of the voltage-gated channel. Thus

$$\delta n(\omega) = \frac{\partial n}{\partial V} (1 - i\omega\tau)^{-1}, \quad (\text{II.C.4.8})$$

so that

$$\delta I(\omega) = gn_o \delta V(\omega) + gV_o \delta n(\omega) \quad (\text{II.C.4.9a})$$

$$= g[n_o + V_o \frac{\partial n}{\partial V} (1 - i\omega\tau)^{-1}] \delta V(\omega). \quad (\text{II.C.4.9b})$$

This corresponds to an effective impedance of

$$\frac{\delta V(\omega)}{\delta I(\omega)} = Z_{eff}(\omega) = g^{-1}[n_o + V_o \frac{\partial n}{\partial V} (1 - i\omega\tau)^{-1}]^{-1}. \quad (\text{II.C.4.10})$$

In parallel with the membrane capacitance  $C$ , this impedance generates a resonance at

$$\omega_o = \left[ \frac{\partial n}{\partial V} (gV_o/\tau C) \right]^{1/2}. \quad (\text{II.C.4.11})$$

The success of the gated channel interpretation of membrane impedances<sup>108</sup> has led to the suggestion that sensory systems incorporate channels which are gated by their stimulus. Thus in electroreceptors we should have the same sorts of voltage-gated channels found in nerve,<sup>36,109,110</sup> while in the inner ear the hair cells should possess channels which are gated by motions of the stereocilia. These channels would act as amplifiers because the opening of a single channel will allow many ions to flow across the cell membrane, which nominally has a standing voltage of  $\approx 0.1$  V. The following arguments are aimed at calculating the noise temperature of this amplifier using the theory of thermal fluctuations in coupled systems.

As in the case of the voltage-gated channel, we imagine  $N$  channels in the hair cell membrane, each of which can exist in two states, and we denote by  $n$  the number of open channels. The analog of Eq. (II.C.4.5) is then

$$\frac{d(\delta n)}{dt} = -(1/\tau)[\delta n - (\frac{\partial n}{\partial x})\delta x] \quad (\text{II.C.4.12})$$

where  $\delta x$  is the displacement of the stereocilium. This describes a channel which "gates" in a time  $\tau$  in response to the stimulus; from the experiments of Corey and Hudspeth<sup>111</sup> we know that  $\tau \approx 40 \mu s$ .

The gated channel system has two generalized coordinates,  $\delta n$  and  $\delta x$ . Correspondingly there must be two generalized forces, of which one is obviously the force applied to the stereocilium; the other generalized force must be taken as the energy difference  $E$  between the two channel states, so that the product of forces and coordinates gives the energy of the system. The equilibrium number of open channels is determined by the energy difference  $E$  through the Boltzmann distribution, so that

$$n_0 = \frac{N}{1 + e^{-E/k_B T}}. \quad (\text{II.C.4.13})$$

We also obtain the "compliance" of the generalized coordinate  $\delta n$  with respect to its generalized force,

$$\frac{\partial n}{\partial E} = \frac{n_0(N - n_0)}{N}. \quad (\text{II.C.4.14})$$

Considering the effects of the generalized force  $\delta E$  on Eq. (II.C.4.12), we obtain upon passing to the Fourier representation

$$\delta n(\omega) = \frac{(\partial n/\partial E)\delta E(\omega) + (\partial n/\partial x)\delta x(\omega)}{1 - i\omega\tau}. \quad (\text{II.C.4.15})$$

Because displacements of the stereocilium change the number of open channels, we expect that changes in the number of open channels should apply a (possibly frequency-dependent) force to the stereocilium. Thus, if the stereocilium were not coupled to the channels we would have

$$\delta x(\omega) = \alpha(\omega)\delta F(\omega), \quad (\text{II.C.4.16})$$

but with coupling we will have

$$\delta x(\omega) = \alpha(\omega)[\delta F(\omega) + \beta(\omega)\delta n(\omega)], \quad (\text{II.C.4.17})$$

where  $\beta(\omega)$  is the "back-coupling" from the channel chemistry to stereocilium mechanics. If we solve Eq's. (II.C.4.15) and (II.C.4.17) for  $\delta n$  and  $\delta x$ , we obtain

$$\delta n(\omega) = \frac{(\partial n/\partial E)\delta E(\omega) + (\partial n/\partial x)\alpha(\omega)\delta F(\omega)}{1 - i\omega\tau - \alpha(\omega)\beta(\omega)(\partial n/\partial x)}, \quad \text{and} \quad (\text{II.C.4.18a})$$

$$\delta x(\omega) = \frac{\alpha(\omega)\beta(\omega)(\partial n/\partial E)\delta E(\omega) + \alpha(\omega)(1 - i\omega\tau)\delta F(\omega)}{1 - i\omega\tau - \alpha(\omega)\beta(\omega)(\partial n/\partial x)}. \quad (\text{II.C.4.18b})$$

These equations have the standard form for two coupled systems, so that  $\beta(\omega)$  is completely determined by the symmetry condition<sup>24</sup> on the response matrix:

$$\beta(\omega) = (\partial n/\partial x)(\partial n/\partial E)^{-1}. \quad (\text{II.C.4.19})$$

Intuitively we know that a signal which is below the thermal noise of the stereocilium cannot deliver much energy to the cilium; in fact this energy will be less than  $k_B T$ . This energy is not sufficient to open very many channels unless the energy difference between open and closed states is very small, in which case spontaneous openings and closings should be rather frequent. This problem similar to that raised by de Vries<sup>112</sup> with regard to the initiation of a nerve impulse, and the methods introduced here make these considerations quantitative.

If we look at the equations for the channel/cilium system in the limit of zero frequency, we find

$$\delta n(\omega = 0) = \frac{\kappa(\partial n/\partial E)\delta E + (\partial n/\partial x)\delta F}{\kappa - \beta(\partial n/\partial x)} \quad (\text{II.C.4.20a})$$

$$\delta x(\omega = 0) = \frac{(\partial n / \partial x) \delta E + \delta F}{\kappa - \beta(\partial n / \partial x)} \quad (\text{II.C.4.20b})$$

where  $\kappa = \alpha^{-1}(0)$  is the stiffness of the stereocilium in the absence of coupling to the channels. The *apparent* stiffness of the cilium is thus

$$\kappa_{eff} = \kappa - \beta(\partial n / \partial x). \quad (\text{II.C.4.21})$$

We must have the stiffness be positive, however, or else the system will be unstable. Therefore

$$\beta(\partial n / \partial x) \leq \kappa, \quad (\text{II.C.4.22})$$

or, from Eq. (II.C.4.19),

$$\frac{\partial n}{\partial x} \leq [\kappa(\partial n / \partial E)]^{1/2}. \quad (\text{II.C.4.23})$$

This equation provides a thermodynamic upper bound on the strength of coupling between the stereocilium and the channels, and hence on the sensitivity of the stimulus-gated channel transducer.

From Eq. (II.C.4.18) we can calculate the spectral density of fluctuations in the number of open channels by applying the fluctuation-dissipation theorem to the coupled system. The general result<sup>24</sup> is that if set of coordinates  $x_i$  respond to forces  $F_i$  such that

$$x_i(\omega) = \sum_j \alpha_{ij}(\omega) F_j(\omega), \quad (\text{II.C.4.24})$$

then the spectral density of the fluctuations in the  $i^{\text{th}}$  coordinate is

$$S_i(\omega) = \frac{k_B T}{\pi \omega} \alpha_{ii}''(\omega), \quad (\text{II.C.4.25})$$

where  $\alpha''$  denotes the imaginary part of  $\alpha$ . To apply this theorem to Eq. (II.C.4.18), we may approximate  $\alpha(\omega) \approx (\kappa - i\omega\gamma)^{-1}$ , where the neglect of stereocilium mass follows from the discussion of Section II.C.3; we also know from that discussion that  $\kappa \gg \gamma\omega$  at normal auditory frequencies. When all the dust clears, the spectral density of fluctuations in the number of open channels is given by

$$S_n \approx (k_B T/\pi) \frac{(\kappa^2\tau)(\partial n/\partial E) + \gamma(\partial n/\partial x)^2}{(\kappa - \beta(\partial n/\partial x))^2}, \quad (\text{II.C.4.26})$$

where I have substituted for  $\beta$  from Eq. (II.C.4.19).

Noise in an amplifier is usually interpreted by referring it to the input. Thus we may view the fluctuations in channel number as being due to an effective level of noise at the stereocilium given by

$$\begin{aligned} S_x^{eff}(\omega) &= (\partial n/\partial x)^{-2} S_n(\omega) \\ &\geq (k_B T/\pi) \frac{\kappa\tau + \gamma}{(\kappa - \beta(\partial n/\partial x))^2} \\ &\geq \frac{k_B T}{\pi} \frac{\gamma}{\kappa^2} (1 + \kappa\tau/\gamma). \end{aligned} \quad (\text{II.C.4.27})$$

Comparing this expression for the effective level of stereocilium noise in the stimulus gated channel with that for the Brownian motion of the isolated cilium [Eq. (II.C.3.1)], we find that the noise in the transducer is greater by a factor  $(1 + \kappa\tau/\gamma)$ . This means that the system is behaving not as if it were at temperature  $T$ , but rather as if it were at the higher temperature  $T' = T + T\kappa\tau/\gamma$ . The difference  $T' - T = T_N$  is called the "noise temperature" of the amplifier.

With the parameters discussed above, the stimulus gated channel has a noise temperature of 12,000 K, or forty times room temperature. This means that the noise

added by the transducer is forty times as large as the noise due to stereocilium Brownian motion. If we expected to detect the threshold signal after it has passed through such a device, we would have to build a filter not of 50 Hz bandwidth (cf. Section II.C.3), but of 1 Hz bandwidth; such filtering is not observed. We conclude that the stimulus-gated channel is inadequate to transduce the threshold auditory signal without some prior amplification. In particular, the channel amplifier cannot explain how the receptor cell is able to reduce thermal noise until the quantum noise becomes limiting. As in the case of the visual system (Section II.A.4) we see that a simple chemical mechanism is incapable of accounting for the quantum-limited performance of biological systems.

The problem with chemical mechanisms is far more serious than even the comparison of noise temperatures implies. It is apparent from the discussion above that by appropriate choices of the parameters it is possible to reduce the noise temperature of the channel system arbitrarily. But the uncertainty principle dictates a minimum noise level independent of the amplification mechanism, so that the prediction of zero noise temperature as a limiting case of the channel model must be in error: the channel model is apparently inconsistent with the uncertainty principle.

The basic principle of any amplifier which is described by chemical kinetics is that the incoming signal modulates the rate(s) of transition(s) among the states of the system. Note that according to this definition a transistor amplifier is a "chemical" device since the incoming voltage signal changes the rates of transition among the various electronic states on either side of the *p-n* junction. In order to understand the apparent violation of the uncertainty principle in such devices, we must look at the microscopic mechanism by which the signal can modulate the rate of a chemical reaction.

Let us focus on one state of the system which has energy  $E_0$  in the absence of the signal; this state may be a chemical state such as “channel open” or a quantum state such as describes the electrons in the transistor. The signal  $x(t)$  is coupled to the amplifier because the energy of the state depends on  $x(t)$ ,

$$E(t) = E_0 + \frac{dE}{dx} x(t). \quad (\text{II.C.4.28})$$

In quantum mechanical terms, the wave function for a system with time varying energy is given by

$$\psi(t) = \exp[-i \int dt' E(t')], \quad (\text{II.B.4.29})$$

so that with a sinusoidal signal  $x = A \sin(\omega t + \phi)$  the wave function becomes the frequency modulated waveform

$$\psi(t) = e^{-iE_0 t} e^{-i\lambda A \cos(\omega t + \phi)}, \quad (\text{II.B.4.30})$$

where  $\lambda = \omega^{-1}(dE/dx)$ . Frequency modulated waveforms can have very different appearances depending on how we observe them.

If the amplifier dynamics is characterized by very strong interactions with the heat bath, so that the relaxation times are very short, then the heat bath “observes” the wavefunction of the amplifier on a very short time scale. As a result, the bath can track the instantaneous frequency of the wavefunction, which implies that the amplifier will equilibrate such that the populations of the states are appropriate to their time-varying energies. In order for this to occur, the rates of transitions among the states must be related to the time variation of the signal, and we recover the standard kinetic description.

On the other hand, if the relaxation times are long the heat bath makes observations only on a much longer time scale, or equivalently samples the wavefunction through a very narrow bandwidth. A frequency modulated waveform, when examined by a narrow bandwidth analyzer, consists of a series of side bands spaced by the modulation frequency. In the case of a wavefunction this means that there are substates of the system which differ in energy by integer multiples of  $\hbar\omega$ , which is just the quantum of energy in the signal. If the amplifier we are considering achieves the quantum limit to its performance, then it must be capable of resolving these substates, and hence of resolving the sidebands in the modulated wavefunction.

Each of the sidebands in the wavefunction has an amplitude and a phase. In order for a chemical kinetic picture to be approximately correct, the phases must be randomized, so that only the populations of the states are relevant. In the modulated wavefunction, however, it may be shown that the phase  $\phi$  of the incoming signal affects only the phase of the wavefunction sidebands, not their amplitudes. Thus in the quantum-limited regime the phase of the signal is translated into the phase of the amplifier wavefunction, as is clearly true in Josephson junction devices, for example. Any system which rapidly randomizes the wavefunction phase will lose information regarding the phase of the signal and cannot function as a linear amplifier. This leads us to two results:

- [1] Kinetic descriptions of amplifier mechanisms are inconsistent with quantum-limited performance for a linear amplifier.
- [2] Quantum-limited linear amplifiers must exhibit a long memory for quantum mechanical phase--they must be *coherent*.

#### D. Some conclusions

Every student of the sensory systems chooses some feature of their behavior at which to marvel. Whether it is our ability to resolve counterpunal themes in a concerto or to perceive the beauty of simple visual patterns, our senses perform remarkable tasks with ease. Even more remarkable is that they have apparently evolved to a sort of "physicist's perfection" that is, each of the systems which we have examined operates at or near to an absolute physical limit to its performance.

It is perhaps not so surprising that the eye can count single photons, since the energy of the absorbed photon is so much larger than  $k_B T$ . We have seen, however, that accounting for this performance is actually quite difficult (cf. Section II.A.4). In the mechano-receptors of the inner ear and electroreceptors of fish, the signal energy is comparable to  $k_B T$ , but quantum limits are nonetheless important. In the electroreceptor the quantum limit arises because the fish makes astonishingly accurate measurements of both amplitude and phase of the signal. In the inner ear, much like the antennae designed to detect gravitational radiation,<sup>27</sup> the signal is entirely classical but is incapable of delivering more than one vibrational quantum to the detector element, so that the driven motions of the detector are comparable to its zero-point motions.

The observation that the threshold signals in the inner ear are comparable to zero-point motion implies--without further argument--that the ear is *not* a classical device. To appreciate the significance of this observation, however, we have to understand the role of amplifiers in quantum-limited measurement, particularly at finite temperature, and we have to understand that quantum noise is not like the more familiar noise sources which can be "processed away" if we know what signal we are looking for and have enough time to find it.<sup>113</sup> What we have seen is that a description of quantum-limited measurement hinges on the existence of a quantum-limited amplifier; in the biological systems this amplifier must be used not only to bring the signal up to the classical level but also to synthesize a filter which can remove the thermal noise.

Very few quantum limited amplifiers have been built in the laboratory, superconducting devices,<sup>25,26,114</sup> masers<sup>115,116</sup> and parametric amplifiers<sup>117</sup> being essentially the only examples. Each of these systems is described by quantum mechanical states in which the individual particles exhibit mutual coherence over macroscopic distances, of order one micron in superconductors and larger in masers. This macroscopic coherence is consistent with the general conclusion derived above, namely that quantum-limited linear amplifiers must preserve phase for times comparable to the measurement. Such coherent behavior is inconsistent with any description of the system in terms of chemical kinetics. We must conclude that *no* kinetic model of events in the sensory systems can account for their observed ability to reach the quantum limits to measurement. Instead these biological systems must be described in explicitly quantum mechanical terms, and their molecular dynamics must be such that quantum coherence is manifest on macroscopic scales of time and distance. To understand these phenomena we must therefore turn to an analysis of the dynamics of biological molecules.

## Chapter Three

### Molecular dynamics and macroscopic implications

*...nature works on a different plan. Her fundamental laws do not govern the world as it appears in our mental picture in any direct way, but instead they control a substratum of which we cannot form a mental picture without introducing irrelevancies....*

*P.A.M. Dirac, 1930<sup>1</sup>*

#### A. A general theoretical approach

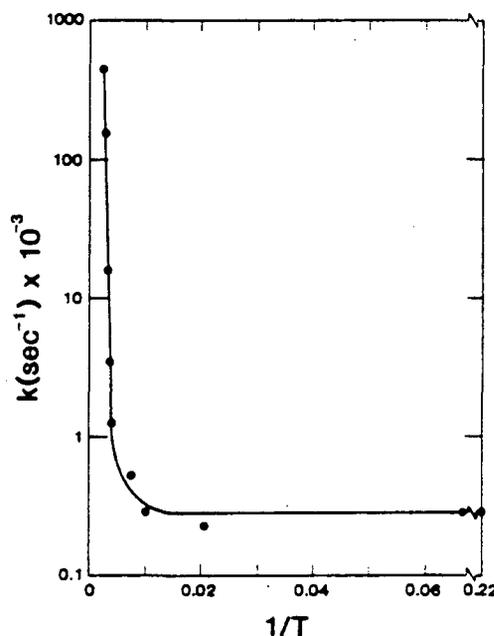
##### 1. *The experimental background*

In developing a theoretical picture of quantum effects in the dynamics of biological molecules we must be guided to some extent by experiment. In particular, several experiments over the last two decades have prompted considerable theoretical interest in the possibility of quantum effects in biology. These experiments will therefore be reviewed not only because they introduce the relevant theoretical issues, but also because they form the body of data against which theory must ultimately be judged.

Historically the key experiments are those of DeVault and Chance,<sup>2,3</sup> who studied electron transfer in the photosynthetic bacterium *Chromatium vinosum*. Photon absorption by this system results in electron transfer from a donor molecule, denoted  $P_{870}$ . Some time after this transfer, the hole on  $P_{870}$  is filled by an electron from a cytochrome *c* molecule. DeVault and Chance studied the kinetics of this latter

reaction, and found, as illustrated in Fig. III-1, that its rate is independent of temperature below 100 K, but exhibits a more typical Arrhenius temperature-dependence above this point. The data may be summarized by three parameters: the low temperature rate  $k(T=0) = 2.8 \times 10^2 s^{-1}$ , the infinite temperature rate  $k(T \rightarrow \infty) = 7.8 \times 10^8 s^{-1}$ , and  $E_a = 0.18 eV$  such that  $k(T \geq 150 K) = k(T \rightarrow \infty) e^{-E_a/k_B T}$ .

FIGURE III-1 Temperature-dependent kinetics of the DeVault-Chance reaction.<sup>3</sup> Note the temperature independence of the reaction rate below  $\approx 100 K$ , compared with the Arrhenius behavior above this temperature.



The temperature independence below 100 K was suggestive of a tunneling process, as had been considered for at least some chemical reactions since the inception of quantum theory.<sup>4</sup> Elementary textbook examples of tunneling<sup>5</sup> involve a quasi-free particle approaching a barrier of height  $\Delta E$  and width  $d$ . The probability of penetrating the barrier is then shown, by directly solving the Schrödinger equation, to be  $P = e^{-d\sqrt{2m\Delta E}/\hbar}$ , where  $m$  is the mass of the tunneling particle. If the particle encounters the barrier  $\nu$  times each second, then we have a temperature independent “reaction rate”

$$k_{\text{tunnel}}(T = 0) = \nu e^{-d\sqrt{2m\Delta E}/\hbar} \quad (\text{III.A.1.1})$$

On the other hand, at high temperatures, we expect the reaction rate to be given by

the usual results of activated complex theory,<sup>6</sup> which means that we identify  $\nu$  as the infinite temperature rate and  $\Delta E$  as the activation energy  $E_a$ . If we assume that it is the transferred electron which tunnels, then  $m = m_e$ , and we obtain  $d \approx 10 \text{ nm}$ .

One of the molecules which participates in the DeVault-Chance reaction (cytochrome *c*) is a member of a well studied class of proteins, and the three-dimensional structures of several members of the class have been solved crystallographically<sup>7</sup> to high spatial resolution. It is known that they are medium size proteins, with diameters on the order of  $1.5 \text{ nm}$ . Thus if the naive tunneling model is correct, the transfer distance is much larger than the sizes of the molecules involved in the transfer--the electron is transferred through a large "space" between donor and acceptor--which is unreasonable.

Fortunately, the naive tunneling picture is itself incorrect, as first shown by Grigorov and Chernavskii.<sup>8</sup> Equations such as (III.A.1.1) are valid only when the densities of states on both sides of the barrier are smooth, so that a particle coming from one side can always find a state of appropriate energy on the other side. But molecular systems have (primarily) bound electronic states with discrete energies. The probability that the energy levels of the two molecules will match is low, and we know that the final state differs from the initial state by nearly half an electron volt. Where does this energy go, and how is the "match" achieved?

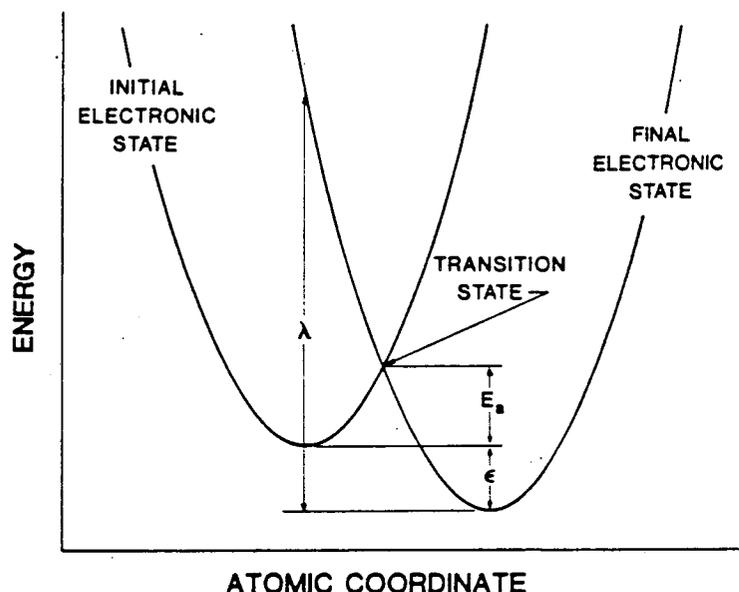
Grigorov and Chernavskii<sup>8</sup> realized that this energy disposal must proceed through the vibrations of the molecules involved in the reaction. This idea is analogous to the description of dissipation of electronic energy in solids: the energy can be lost to collisions with other electrons (not relevant to a simple donor/acceptor pair), to photon emission (not observed), or to phonon emission. Because the phonons, or vibrational quanta, obey Bose-Einstein statistics, phonon emission can be either spontaneous or stimulated, as in the case of photons. At zero temperature there are no phonons present so that only the spontaneous process occurs, and this occurs at a

temperature-independent rate until the temperature is sufficiently high that the mean number of phonons is greater than one. Once this occurs, stimulated phonon emission takes over, and the rate of phonon emission continues to rise with temperature.

This basic picture of the temperature dependence reflecting a cross-over from spontaneous to stimulated phonon emission is very different from the initial picture of electron tunneling. We have not made clear, however, how the electron transfer is coupled to the phonon emission. The nature of this coupling, together with the frequencies and linewidths of phonons, will determine the scale of the transfer rate, although the interpretation of its temperature dependence in terms of phonon populations will be seen to be correct in light of more rigorous theories.

Hopfield<sup>9</sup> made use of a semi-classical theory to describe the coupling between electron and phonon degrees of freedom. The results of his analysis may be understood in terms of a configuration-coordinate diagram such as Fig. III-2. In each of the two relevant electronic states we trace out the potential energy of the system as a function of some generalized coordinate (configuration) describing the positions of the atoms in the molecule. The equilibrium positions of the atoms and the vibrational frequencies about these positions will be different in the two states, and there will be an energy gap between the states. The reaction we are trying to understand is a jump from one surface to the other under the influence of some matrix element  $V$ .

Conservation of energy demands that the jump occur horizontally on the diagram of Fig. III-2. The Franck-Condon principle demands that the jump occur vertically, since the time scale for vibrational motion is much slower than that for electronic motion.<sup>10</sup> These two requirements can be met simultaneously at only one point; namely at the intersection between the two surfaces. The rate of the reaction is therefore proportional to the probability of being in the neighborhood of this point, as well as to the square of the matrix element  $V$ , which follows from perturbation theory (Fermi's golden rule).<sup>5</sup> If we treat the molecular vibrations as simple harmonic



**FIGURE III-2** Semi-classical approach to electronic transitions. The problem is to calculate the transition rate from the upper to the lower surface, which differ by an energy gap  $\epsilon$  and a "reorganization energy"  $\lambda$ . Application of the Franck-Condon principle and a semi-classical approximation lead to Eq. (III.A.1.2), as described in the text.

oscillators, then the probability distributions for the coordinates are Gaussian with variances proportional to  $T_{eff} = \hbar\omega \coth(\hbar\omega/2k_B T)$ , where  $\omega$  is the vibrational frequency; note that  $T_{eff} \rightarrow T$  at large  $T$  and  $T_{eff} \rightarrow \frac{1}{2}\hbar\omega$  at small  $T$ . The probability of being at the intersection point is then easily calculated, and the rate constant becomes

$$W = (\text{constant}) |V|^2 e^{-(\epsilon - \lambda)^2 / 2\lambda T_{eff}}, \quad (\text{III.A.1.2})$$

where  $\epsilon$  is the energy gap and  $\lambda$  is the reorganization energy defined in Fig. III-2.

Two points should be noted about this semi-classical result. First, at large  $T$  we obtain  $W \approx W(\infty) e^{-E_a/k_B T}$ , where  $E_a$  is an apparent activation energy equal to the energy required to travel up to the intersection point. This prediction is very similar to the transition state, or activated complex view of chemical reactions.<sup>6</sup> We thus identify the transition state picture as a semi-classical one, and this is consistent with modern work on gas-phase reaction theory.<sup>11</sup>

Almost all biochemical processes have been interpreted in terms of transition state theory, and Pauling used this theory in his application of the structure-function principle to enzymatic catalysis.<sup>12</sup> He suggested that an enzyme has a structure complementary to the transition state of the reaction it must catalyze. As a result, the greatest stabilization of the enzyme-substrate complex would occur in the transition state, lowering the energy of the transition state relative to that of the reactants. This selective lowering of the transition state energy could then account for the acceleration of the reaction rate by the enzyme.

Transition state stabilization remains the central feature of all modern discussions of catalytic mechanisms in biochemistry.<sup>13,14</sup> We see that this concept depends on two approximations: semi-classical behavior of the molecule and the high-temperature limit. Analysis of the DeVault-Chance and related reactions therefore gives us an opportunity to test the approximations which form the basis of one of the most widely applied theories in biochemistry.

The second point about the semi-classical rate expression is that it is finite in the low-temperature limit. This arises from the possibility that the zero-point motion of the molecule will carry it to the transition state. The probability distribution for the zero-point motion extends outside the region which would be allowed to a classical particle with the zero-point energy, as shown in Fig. III-3. Thus the finite reaction rate at zero temperature is a manifestation of quantum mechanical tunneling of the *vibrations* rather than the electron.

Jortner<sup>15</sup> attempted to go beyond Hopfield's semi-classical treatment by borrowing the theory of radiationless transitions which had been developed to describe F-centers and other impurities in solid state systems.<sup>16,17</sup> The basic idea is to fill in the quantum mechanical states of the molecular vibrations, as shown in Fig. III-4, and then apply the golden rule to calculate transition rates among particular pairs of electronic/vibrational ("vibronic") states; the final rate expression is then obtained by

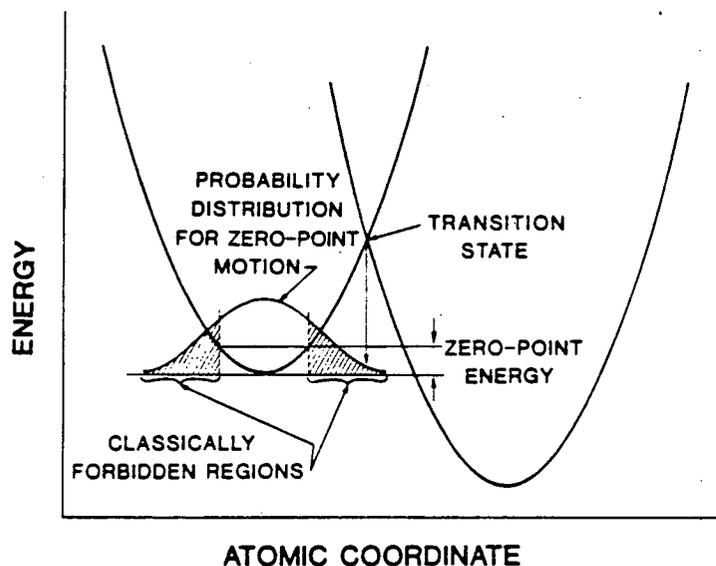


FIGURE III-3 Vibrational tunneling and low-temperature reaction rates. Same problem as Fig. III-2, where the transition from one electronic state to the other can occur only at the intersection between the potential surfaces. At absolute zero, the probability of reaching this intersection point is determined by the statistics of zero-point motion, whose probability distribution is shown. A classical particle with energy equal to the zero-point energy could not move into the "classically inaccessible regions," which include the intersection point, so that the reaction occurs only as a result of atomic tunneling.

a thermal average over the initial vibrational states. In solid state systems this works quite well, since the vibrational modes of the lattice form a continuum so that there is always a nearly degenerate state into which the system can make a transition. Molecular systems, however, have discrete vibrational modes, and it is only through interactions between the molecule and the solvent--vibrational relaxation--that a continuum can be generated.

Jortner did not give a rigorous treatment of vibrational relaxation, and instead argued for "coarse-graining" on a scale fixed by typical frequencies of solvent modes.<sup>15</sup> His rate expression in turn depends on this ill-defined frequency, which hampers any attempt to rigorously test the theory; without coarse-graining the transition probability diverges. In addition, Jortner's approach *via* vibronic state to vibronic state transitions is hard to generalize to more realistic Hamiltonians. An alternative approach, developed by Soules and Duke<sup>18</sup> to treat energy transfer among impurities

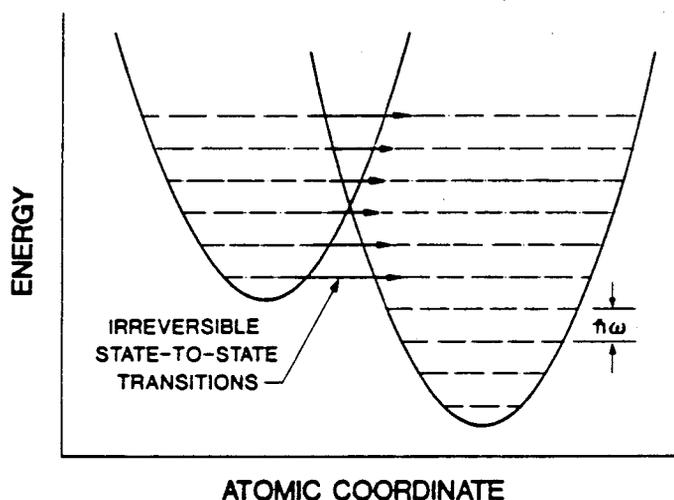


FIGURE III-4 "State-to state" transition analysis. The transition from one electronic state to the other is broken into the individual vibronic transitions, each of which is assigned a rate calculated from the Golden rule.

in solids, writes the rate of electronic transitions as the spectral density of some unitary operator which defines the transformation from the potential surface of the reactants to that of the products. This method and variants upon it have been further developed by a number of investigators.<sup>19-23</sup>

None of these methods, however, take explicit account of vibrational relaxation. They would all predict infinitely sharp Raman or infrared spectra for the systems described, and all require some external postulates to guarantee irreversibility of the transition and convergence of the integrals which define the transition probability. For example, Kenkre, Knox and co-workers<sup>24,25</sup> make use of a coarse-graining procedure, while Silbey and co-workers<sup>20-23</sup> restrict their calculations to systems in which the vibrational modes form a natural continuum.

In summary, qualitative consideration of the DeVault-Chance data suggests that the initial descriptions of electron tunneling are misleading. Instead we focus upon the coupling between electronic states and vibrational dynamics in the molecule, either intuitively in terms of spontaneous and stimulated phonon emission or semi-classically

in terms of the configuration-coordinate diagram. Both approaches lead to a partial understanding of the temperature dependence of the reaction rate, and the semi-classical method supports the applicability of transition state theory as a high temperature approximation. To go beyond the semi-classical approximation, and hence test for true quantum effects, we must however take rigorous account of vibrational relaxation in the molecule; existing theoretical methods cannot do this.

Many of the same theoretical issues which arise in the case of the DeVault-Chance reaction are also significant in the case of the heme proteins and their interactions with small ligands. Myoglobin (Mb), for example, binds carbon monoxide (CO) by a covalent link to the active site iron atom; this bond can be photodissociated within one picosecond<sup>26</sup> and with quantum yields near unity. As time passes, the photodissociated ligand can make its way out of the protein matrix into the solvent or find its way back to the iron atom. The reappearance of the ligand at the iron atom can be monitored by changes in optical absorption which occur upon ligation.

The rebinding reaction can be resolved into several stages.<sup>27</sup> One involves crossing the barrier that divides the interior of the protein from the bulk solution, and all other stages are true intramolecular processes. When the system is cooled below the freezing point of the solvent, the ligand cannot escape the protein interior, so that only the intramolecular processes are observed. These processes occur with a non-exponential time course; in fact, as illustrated in Fig. III-5, the time course is well approximated by a power-law decay  $n(t) \approx t^{-\nu}$  over several orders of magnitude in time.<sup>27</sup> By choosing some representative point, *e.g.*  $n = 0.75$ , it is possible to define the effective rate  $t_{0.75}^{-1}$  of the process, and to examine the dependence of this rate on temperature. The data exhibit the same pattern as observed by DeVault and Chance--the rate is temperature independent below approximately 20 K and shows classical Arrhenius behavior above this temperature.<sup>28-31</sup>

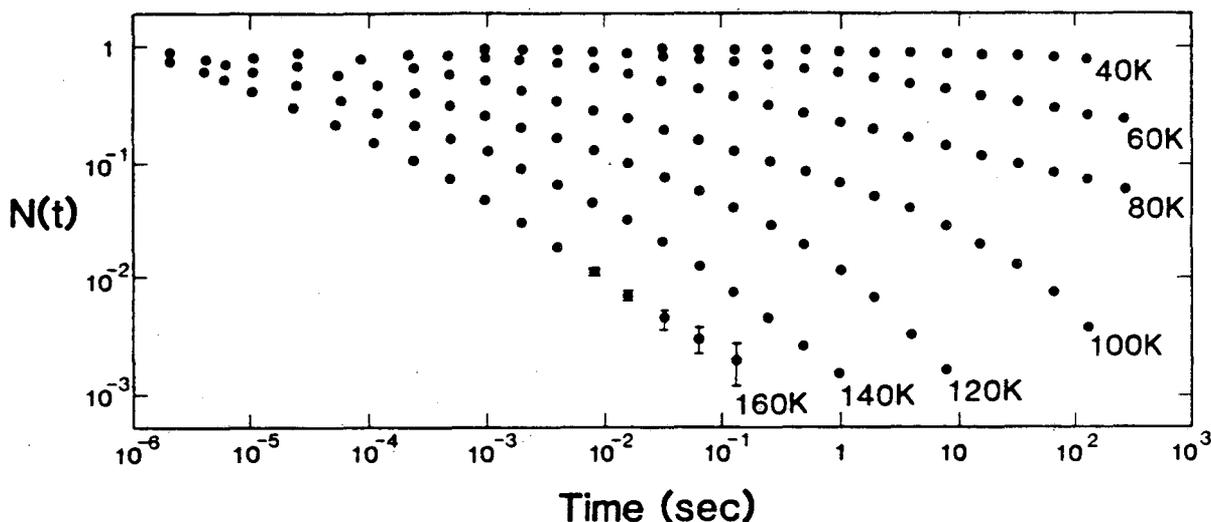
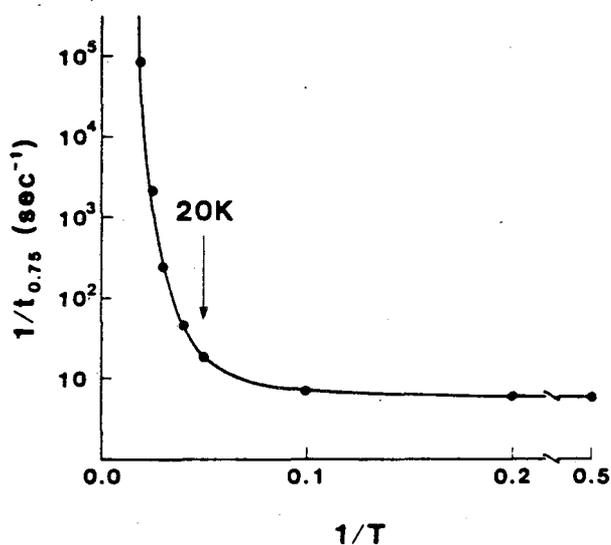


FIGURE III-5 Kinetics of ligand binding in heme proteins. Above, the time course of rebinding of CO to Mb, shown on a double-logarithmic scale to emphasize the nearly power-law behavior over several orders of magnitude in time (after Ref. 27). Right, temperature dependence of the "average rate," measured as the inverse of the time  $t_{0.75}$  required for one-quarter of the molecules to rebind; data in this case refer to CO binding to isolated chains of hemoglobin,<sup>28,30,31</sup> but the pattern is similar in myoglobin.



The non-exponential time course of the reaction may be described as arising from a distribution of reaction rates over the ensemble of molecules studied in the experiment. This suggests that not all molecules of the sample are identical, and that the variations in reaction rate which result from the differences among molecules can amount to several orders of magnitude. Clearly the intermolecular differences are subtle, and the challenge is to identify the critical variables which characterize these differences and then explain why the rate constant is so sensitive to these parameters.

Recent experiments have shown that non-exponential decay at low temperatures can also be observed (under some conditions) for the charge-recombination reactions

of photosynthesis,<sup>32</sup> suggesting that myoglobin may not be so unusual. It remains to be seen whether comparably accurate experiments in other biological systems will provide further evidence of this effect.

The heme protein reactions extend the temperature-dependence problem, which was first noticed in the DeVault-Chance reaction, by requiring not only a single reaction rate with non-classical temperature dependence but a whole distribution of such rates. Another direction in which the theoretical issues can be extended is by consideration of reactions in the picosecond regime, where they can compete with the processes of vibrational relaxation. On this time scale we expect that the neglect of vibrational relaxation in current theories will be particularly serious.

The advent of picosecond and sub-picosecond spectroscopies has allowed the primary events of a number of light induced biological processes to be directly observed.<sup>33,34</sup> In particular, several intermediates in bacterial photosynthesis have been detected,<sup>35</sup> and in some cases the absorption spectra of these intermediates can be compared with those of species which have been trapped by chemical manipulations.<sup>36</sup>

Figure III-6 shows the basic scheme for the primary processes of photosynthesis in the purple bacterium *Rhodospirillum rubrum*. At least two reactions--electron transfer out of the initial photoexcited state  $P^*$  to the intermediate  $I$  and the subsequent transfer from  $I$  to  $Q_A$ --fall in the picosecond domain. Both these reactions have a peculiar temperature dependence. The  $P^*I \rightarrow P^+I^-$  transition<sup>37</sup> occurs in about three picoseconds at helium temperatures and slows down by a factor of two as the temperature is raised to 300 K. The transition  $I^-Q_A \rightarrow IQ_A^-$  has a non-monotonic temperature dependence with a reaction rate maximum at 25 K; this non-monotonicity disappears if the solvent is deuterated.<sup>38</sup> Thus the picosecond reactions, while exhibiting large regions of temperature independent reaction rates which might be interpreted as tunneling, do not exhibit the simple pattern of the DeVault-Chance reaction.

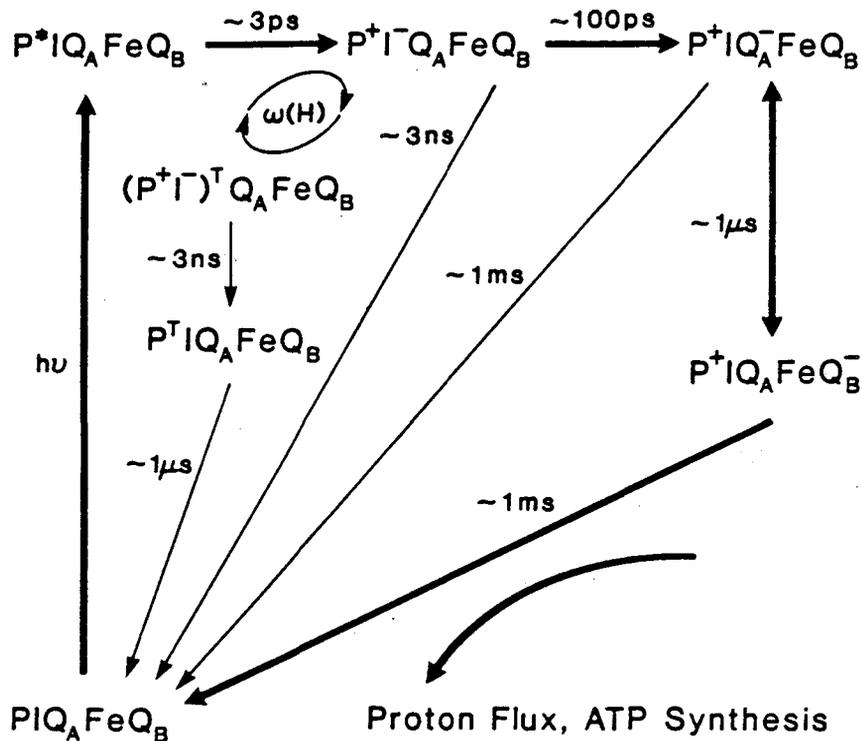


FIGURE III-6 Primary events of bacterial photosynthesis. This figure schematizes the light-induced electron transfer reactions which occur within the "reaction center."  $P$ ,  $I$ ,  $Q$ , etc. are all molecules bound in the reaction center which have been positively identified as electron carriers:  $P$  is a "pigment," most likely a dimer of bacteriochlorophylls;  $I$  is an "intermediate" electron acceptor, identified from optical and paramagnetic resonance spectra as a bacteriopheophytin;  $Q_A$  and  $Q_B$  are quinone molecules which interact magnetically with a single iron atom ( $Fe$ ). Asterisk denotes an electronically excited state reached by photon ( $h\nu$ ) absorption, while  $+$  and  $-$  denote cation and anion radical species, respectively;  $T$  denotes a triplet state for those states with two interacting unpaired spins. Time scales for transitions among the states of the system are given to an order of magnitude, and the coherent mixing among singlet and triplet radical pair states ( $\omega(H)$ ) is dependent on the magnetic field strength, as indicated. Under physiological conditions the cycle of electron transfer in the reaction center is coupled to other chemical reactions, as shown, although these do not occur within the reaction center itself.

In terms of configuration coordinate diagrams, it has been suggested<sup>39</sup> that very fast reactions may be understood as shown in Fig. III-7. For special values of the energy gap and configuration change from reactants to products, there is no activation energy for the forward reaction since the potential surfaces cross at the equilibrium position of the molecule in the reactant state. Furthermore, as the temperature increases, the molecule spends less time near its equilibrium position so that the

reaction rate should decrease, as observed. This interpretation assumes, however, that the equilibrium distribution of molecular coordinates is achieved sufficiently rapidly to be relevant in the reaction. But some chlorophyll molecules require at least a few picoseconds to come to vibrational equilibrium,<sup>40</sup> so that the observed time scales for the reaction and vibrational relaxation are comparable, and the interpretation of the kinetics becomes clouded.

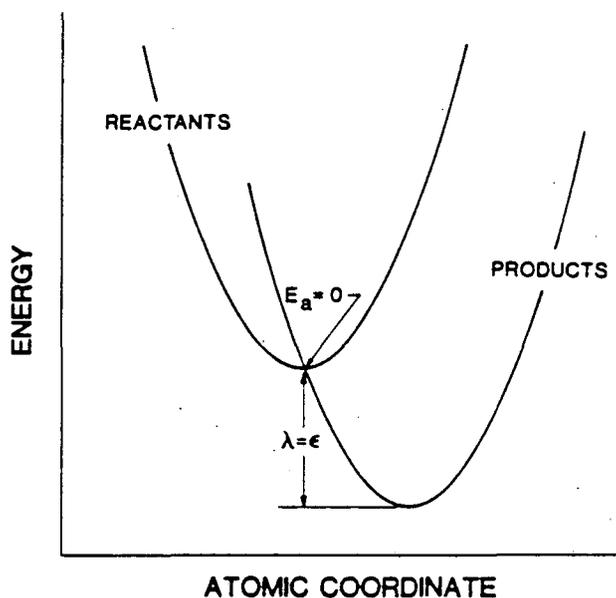


FIGURE III-7 Semi-classical analysis of picosecond transitions. This is the same as in Fig. III-2, but for a special choice of the reorganization energy  $\lambda$ . In this case, there is no "barrier" in reaching the point where the surfaces intersect, and the reaction rate thus has the maximum value consistent with the electronic matrix element.

Almost all problems in condensed matter physics involve interactions among degrees of freedom with well separated relaxation times. When this is not the case, the theory becomes very difficult, because the dynamics of both sets of modes must be considered explicitly. In the present case, it is apparent that the equilibration of the electronic degree of freedom is not well separated in time from that of the vibrational degree of freedom. Thus a rigorous theory must be based on a careful treatment not only of electron-phonon interactions, but also of the vibrational relaxation processes which ultimately account for the approach of the system to thermal equilibrium. As noted above, such a treatment has not yet been given.

The problems of the picosecond regime may be more generally relevant, since some modes of vibration in biological macromolecules are observed to live for nanoseconds and longer,<sup>41</sup> suggesting that even "normal" reactions might not be well described by equilibrium theories. In particular, we shall see that in the non-equilibrium regime there is no true irreversible chemical reaction, essentially because the molecules cannot "know" to what temperature they should be equilibrating.

In summary, those biological systems which have been studied over a wide range of temperatures and time scales do not exhibit the classical Arrhenius behavior across the full temperature range of the study. In particular, some exhibit regions of temperature independence in the reaction rate. These temperature independent rates have been widely interpreted as evidence for quantum tunneling contributions to the reaction rate.

The theories on which these observations have been interpreted are not overly rigorous, and do not allow us to answer unambiguously the question of whether or not quantum effects are important. Theory fails completely in dealing with the picosecond reactions, which compete with the processes of vibrational relaxation, and thus can be understood only by rigorous treatment of the relaxation process. Rigorous treatment of vibrational relaxation is also required to go beyond the semi-classical approximation for the slower reactions, and the following sections provide such a treatment. I begin by developing a family of model Hamiltonians which can describe the dynamics of biological molecules, and then discuss the properties of these Hamiltonians and their predictions in both slow and fast regimes.

## *2. Model Hamiltonians for biological molecules*

Each of the candidate systems for the observation of microscopic quantum effects in biology are, in a certain sense, rather simple: each has only a few relevant electronic states. In the electron transfer reactions of photosynthesis these few states are obvious--reactants and products correspond to electrons localized on physically distinct

sites. In myoglobin, the iron converts from high spin to low spin upon ligand binding, so that the change in electronic state consists of a rearrangement of the populations of the iron  $d$  orbitals.

Ideally we may imagine an experiment which probes the electronic state--and only the electronic state--of the system. In quantum mechanical terms this would mean that our instrument corresponds to a measurement operator whose eigenvalues  $\lambda_n$  label the electronic states  $|n\rangle$ . In practice, as we shall see in the following sections, this is seldom the case in the experiment of interest.

In general a measurement operator cannot encompass all of the degrees of freedom in a system. In particular, while we distinguish among electronic states, the states of motion of the nuclei have not been described. Thus if we write the energy of the system in each electronic state, it will be a function of the coordinates and momenta of all the nuclei in the molecule and the surrounding solvent. This consideration leads us to the general Hamiltonian

$$\mathbf{H} = \sum_i \mathbf{H}_i(\bar{p}, \bar{x}) c_i^\dagger c_i + \sum_{ij} V_{ij}(\bar{p}, \bar{x}) c_i^\dagger c_j. \quad (\text{III.A.2.1})$$

The operators  $c_i^\dagger$  ( $c_i$ ) create (annihilate) molecules in the electronic state  $i$ , and hence obey the fermion anti-commutation relations

$$c_i^\dagger c_j + c_j c_i^\dagger = \delta_{ij}, \quad (\text{III.A.2.2})$$

while  $\bar{p}$  and  $\bar{x}$  denote the momenta and positions on the nuclei, respectively. The  $\mathbf{H}_i$  define the energy of the nuclear motions in a fixed electronic state  $i$ , and the matrix elements  $V_{ij}$  define the amplitudes for transitions from  $j$  to  $i$ .

The Hamiltonian of Eq. (III.A.2.1) contains no approximations; it is the most general model consistent with the enumeration of a countable set of electronic states. It is only when we make approximations to the terms in this Hamiltonian that we lose the exactness of the theory. To the extent that the forms of these terms can be

derived from independent experiments the theory can be made as detailed as the data allow. What we hope is that some physically motivated approximations will be borne out by such independent experiments.

The first approximation is the division of the nuclear coordinates into two classes, the "quantum modes" and the "heat bath." The heat bath is defined as a set of degrees of freedom with the (collective) properties of infinite heat capacity and zero relaxation time to equilibrium. As a result the heat bath remains in thermodynamic equilibrium through any change in state of the "proper subsystem," namely the electronic states and the quantum modes, and in fact the heat bath serves to define the temperature of the system. The quantum modes are coupled to the heat bath, and this coupling allows them to relax to thermal equilibrium, but this relaxation process will occur only at some finite rate. An obvious candidate for the heat bath in a biological system is the aqueous solvent or membrane matrix.

The utility of the heat bath--quantum mode separation stems from the work of Senitzky<sup>42,43</sup> and others,<sup>44,45</sup> who showed that the coupling of a quantum system to a heat bath provides a consistent way of describing the thermal equilibration of the quantum system. For example, we imagine a simple harmonic oscillator coupled to a heat bath through its momentum. The Hamiltonian is

$$\mathbf{H} = \frac{1}{2}\hbar\omega(P^2 + Q^2) + P\Gamma + \mathbf{H}_{hb}, \quad (\text{III.A.2.3})$$

where  $P$  and  $Q$  are the (dimensionless) momentum and position of the oscillator, respectively,  $\Gamma$  is some generalized coordinate of the heat bath, and  $\mathbf{H}_{hb}$  is the Hamiltonian of the heat bath itself. Since the heat bath is assumed to have a large heat capacity and short relaxation time, the perturbation  $P\Gamma$  will have a larger effect on the oscillator than on the heat bath; this suggests doing a perturbation calculation to second order for the oscillator but to first order for the bath. This seemingly inconsistent application of perturbation theory leads to very sensible results which, for less

general models, can be also obtained by more rigorous means. These results<sup>45</sup> are that the creation and annihilation operators  $a = (P + iQ)/\sqrt{2}$  and  $a^\dagger = (P - iQ)/\sqrt{2}$  obey the "quantum Langevin equations,"

$$i\frac{da}{dt} = \omega a - i\gamma a + \delta F \quad (\text{III.A.2.4a})$$

$$i\frac{da^\dagger}{dt} = -\omega a^\dagger - i\gamma a^\dagger + \delta F^\dagger, \quad (\text{III.A.2.4b})$$

where the operators  $\delta F$  and  $\delta F^\dagger$  have the properties

$$\int d\tau e^{i\Omega\tau} \langle \delta F^\dagger(t) \delta F(t-\tau) \rangle = 2\gamma \bar{\nu}(\Omega) \quad (\text{III.A.2.5a})$$

$$\int d\tau e^{i\Omega\tau} \langle \delta F(t) \delta F^\dagger(t-\tau) \rangle = 2\gamma (\bar{\nu}(\Omega) + 1) \quad (\text{III.A.2.5b})$$

$$\langle \delta F(t) \delta F(t') \rangle = \langle \delta F^\dagger(t) \delta F^\dagger(t') \rangle = 0, \quad (\text{III.A.2.5c})$$

and  $\bar{\nu}(\Omega) = [e^{\hbar\Omega/k_B T} - 1]^{-1}$ , where  $T$  is the temperature of the heat bath. In these equations  $\langle \dots \rangle$  denotes an average over the degrees of freedom in the heat bath, a "thermal average." From these equations it may be shown that the average energy of the oscillator relaxes exponentially toward its equilibrium value (with time constant  $1/2\gamma^{-1}$ ), and that the average equal-time commutator  $\langle [a^\dagger(t), a(t)] \rangle$  is a constant of the motion, so that this description of dissipation is consistent with the uncertainty principle.<sup>46</sup>

It may also be shown that the effective Langevin forces are Gaussian, in that their correlation functions always factor into products of the two-point correlation functions of Eq's. (III.A.2.5). Because Eq's. (III.A.2.4) are linear, this factorization is also true for the creation and annihilation operators of the oscillator. In field-theoretic terms, this means that the many-particle Green's functions of the oscillator can be written as products of the single-particle Green's function--a form of Wick's theorem

is valid. Thus the "heat bath method" for describing dissipation in a harmonic oscillator is quite simple: It results in simple exponential decays of the creation and annihilation operators, and hence of the system energy, and preserves Wick's theorem. Therefore if the quantities of interest are reduced to functions of the oscillator Green's function, we simply replace the undamped zero-order function

$$D^{(0)}(\Omega) = i \int d\tau e^{i\Omega\tau} \langle T[a^\dagger(t-\tau)a(t)] \rangle = \frac{1}{\omega - \Omega + i\eta} \quad (\text{III.A.2.6})$$

by the damped version

$$D^{(0)}(\Omega) = \frac{1}{\omega - \Omega + i\gamma}, \quad (\text{III.A.2.7})$$

where  $T$  denotes time ordering,  $\eta \rightarrow 0^+$ , and for simplicity I consider the case of zero temperature. *All* of the effects of the many heat bath coordinates have now been summarized in a single measurable parameter  $\gamma$ . By explicitly considering damped oscillators from the beginning, we can avoid a number of problems in demonstrating that the coupled electron-phonon system comes to thermal equilibrium.

Once we have singled out the coordinates of the quantum modes we can proceed quite directly. The Hamiltonian of the quantum modes in a particular electronic state may be written as a sum of kinetic and potential energy terms, the latter describing the adiabatic (in the Born-Oppenheimer sense) energy surface on which the system moves. This surface will have some absolute minimum, corresponding to the position of the  $\mu^{\text{th}}$  atom  $q_\mu = \Delta_{\mu i}$  in the  $i^{\text{th}}$  electronic state. Expanding the potential energy around this minimum, the lowest order (quadratic) terms allow the definition of the usual vibrational normal modes, with the possibility that the frequencies and structures of these modes depend upon the electronic state.

The state-dependent normal modes may be described as follows. Ordinarily the annihilation operator for quanta (phonons) of the  $\nu^{\text{th}}$  mode is given by (with  $\hbar = 1$

throughout)

$$a_\nu = \sum_\mu \left\{ A_{\nu\mu} (q_\mu - q_\mu^{(0)}) \omega_\nu^{1/2} + i A_{\nu\mu}^{-1} p_\mu \omega_\nu^{-1/2} \right\}, \quad (\text{III.A.2.8})$$

where  $q_\mu^{(0)}$  is the equilibrium position of the  $\mu^{\text{th}}$  atom,  $\omega_\nu$  is the frequency of the  $\nu^{\text{th}}$  normal mode, and the matrix  $A_{\nu\mu}$  expresses the particular combinations of atomic coordinates which constitute the normal modes. To take account of the state-dependence of each of these quantities, we introduce a new annihilation operator

$$\bar{a}_\nu = \sum_\mu \left\{ A_{\nu\mu} (\{c_i^\dagger c_i\}) (q_\mu - \bar{\Delta}_\mu) \bar{\omega}_\nu^{1/2} + i A_{\nu\mu}^{-1} (\{c_i^\dagger c_i\}) p_\mu \bar{\omega}_\nu^{-1/2} \right\}, \quad (\text{III.A.2.9})$$

$$\bar{\Delta}_\mu = \sum_i \Delta_{\mu i} c_i^\dagger c_i \quad \bar{\omega}_\nu = \bar{\omega}_\nu + \sum_i \rho_{\nu i} c_i^\dagger c_i, \quad (\text{III.A.2.10})$$

where  $\bar{\omega}_\nu$  is the average frequency of the  $\nu^{\text{th}}$  mode, and  $\rho_{\nu i}$  is the frequency shift of this mode in the  $i^{\text{th}}$  electronic state; the explicit dependence of the normal mode matrix  $A_{\nu\mu}$  on the electronic state is of the general form

$$A_{\nu\mu} (\{c_i^\dagger c_i\}) = B_{\nu\mu} + \sum_i C_{\nu\mu}^i c_i^\dagger c_i. \quad (\text{III.A.2.11})$$

Using these operators, we may write

$$\sum_i \mathbf{H}_i c_i^\dagger c_i \approx \sum_i \epsilon_i c_i^\dagger c_i + \mathbf{H}_{vib}, \quad (\text{III.A.2.12})$$

where the vibrational Hamiltonian is

$$\mathbf{H}_{vib} = \sum_\nu \bar{\omega}_\nu (\bar{a}_\nu^\dagger \bar{a}_\nu + 1/2) + \text{damping}, \quad (\text{III.A.2.13})$$

and *damping* denotes those terms involving the heat bath.

The fact that the creation and annihilation operators for vibrational quanta in the molecule depend upon the electronic state implies a form of electron-phonon coupling;

this formalizes the notion that the atoms will try to readjust to new equilibrium positions and normal mode frequencies when the electronic state is changed. This coupling allows for energy to flow from the electronic states into the vibrational states, and this energy transfer is completely reversible. Irreversibility is obtained only when we couple the vibrational system to the heat bath and allow for vibrational relaxation.

To understand the effects of coupling between electronic and vibrational degrees of freedom in quantitative terms, we must estimate the strength of this coupling. One measure of the coupling strength is provided by the magnitude of the change in equilibrium position of an atom when the electronic state is changed,  $\xi_{ij}(\mu) = \Delta_{\mu i} - \Delta_{\mu j}$ . To convert this change into a dimensionless coupling constant, we must compare it to the natural length scale given by the zero-point motion of the atom.

Consider a segment of alpha-helical protein of length  $L$ . This structure can vibrate in a number of modes, of which we focus upon the lowest frequency uniform stretch. The properties of this mode can be understood<sup>47</sup> by taking the helix as a uniform elastic rod of area  $A = 0.5 \text{ nm}^2$ , density  $\rho = 1.5 \times 10^3 \text{ kg-m}^{-3}$ , and Young's modulus  $Y = 2 \times 10^{10} \text{ Nt-m}^{-2}$ . The stretching motion is resisted by a stiffness  $\kappa = YA/L$ , and involves the motion of a mass  $M \approx \rho AL$ , so that the vibrational frequency is

$$\omega = (\kappa/M)^{1/2} = (Y/\rho)^{1/2} L^{-1}, \quad (\text{III.A.2.14})$$

or, in conventional units,

$$\omega = 20 \text{ cm}^{-1} (L/1 \text{ nm})^{-1}. \quad (\text{III.A.2.15})$$

The (root-mean-square) zero-point motion is given by

$$\delta x_{z.p.} = (\hbar/2M\omega)^{1/2} = 4 \times 10^{-12} \text{ m}, \quad (\text{III.A.2.16})$$

independent of helix length. Thus, since helix segments in proteins generally run

from a few to ten angstroms, we are dealing with vibrational frequencies from 20 to  $\approx 100 \text{ cm}^{-1}$ . The changes in equilibrium positions of atoms observed crystallographically are on the order of one-tenth angstrom, or  $\approx 2.5$  times the zero-point motion. We shall see that the appropriate dimensionless coupling constant is the square of the displacement ratio, so that we expect coupling constants of about five for these low-frequency modes.

The electron-phonon coupling constants in biological molecules are thus well outside the region where we expect perturbation theory to be successful, namely small relative to one. Furthermore, the modes to which large coupling is expected are in an interesting frequency range: as we raise the temperature above  $T_c = \hbar\omega/2k_B$  we begin to populate the excited vibrational states, while below this temperature we are trapped in the vibrational ground state. For the modes of interest here, the critical temperatures will range from 15 to 100 K, which corresponds well with the ranges over which unusual temperature dependences are observed for biological rate processes. Thus it is possible that the electron-phonon coupling mediates these unusual temperature dependencies, and conversely that the observed temperature dependences probe the effects of the vibrational spectrum on the reaction rate.

What we have done is to identify a class of relatively simple model Hamiltonians which may describe the dynamics of biological macromolecules. Already we see, however, that the physics of these Hamiltonians is quite rich, and can only be understood if we respect certain physical conditions, such as the non-perturbative character of the electron-phonon coupling. The following sections are devoted to the development of calculational methods which allow us to extract the interesting features of these models.

### *3. Calculational methods*

To make progress in the analysis of the Hamiltonian in Eq. (III.A.2.1), it is useful to transform from state-dependent phonon operators to "true" phonon operators.

The "true" phonon Hamiltonian may be described as the Hamiltonian which governs the equilibrium state to which the phonons attempt to relax; clearly this state depends upon the average electronic state of the system, but cannot depend upon the detailed fluctuations in electronic state. If for example the transitions among electronic states were slow, and the model Hamiltonian is that of Eq. (III.A.2.12), then it is apparent that the phonons would be equilibrating with an adiabatically varying Hamiltonian given by

$$\mathbf{H}_{ph} = \sum_{\nu} \omega_{\nu} (a_{\nu}^{\dagger} a_{\nu} + 1/2) + \text{damping}, \quad (\text{III.A.3.1})$$

where  $a_{\nu}$  is obtained from  $\bar{a}_{\nu}$  by replacing  $c_i^{\dagger} c_i$  with  $\langle c_i^{\dagger} c_i \rangle$ , and similarly with  $\omega$  from  $\bar{\omega}$ . The operators  $a_{\nu}$  thus have simple properties, since their zero-order Hamiltonian is decoupled from the electronic degrees of freedom. The problem is to construct the transformation from  $\bar{a}$  to  $a$  and to define its action on the full model Hamiltonian, which is

$$\mathbf{H} = \sum_i \epsilon_i c_i^{\dagger} c_i + \sum_{\nu} \bar{\omega}_{\nu} (\bar{a}_{\nu}^{\dagger} \bar{a}_{\nu} + 1/2) + \text{damping} + \sum_{ij} \bar{V}_{ij} (\bar{a}_{\mu}^{\dagger} \bar{a}_{\mu}^{\dagger}) c_i^{\dagger} c_j. \quad (\text{III.A.3.2})$$

The unitary operator  $U$  such that  $U \bar{a}_{\nu} U^{\dagger} = a_{\nu}$  will commute with  $c_i^{\dagger} c_i$  but not with  $c_i^{\dagger} c_{j \neq i}$ , so that the transformed Hamiltonian may be written

$$\begin{aligned} U \mathbf{H} U^{\dagger} = & \sum_i [\epsilon_i + D_i(a_{\nu}; a_{\nu}^{\dagger})] c_i^{\dagger} c_i + \sum_{\nu} \omega_{\nu} (a_{\nu}^{\dagger} a_{\nu} + 1/2) + \text{damping} \\ & + \sum_{ij} V_{ij}(a_{\nu}; a_{\nu}^{\dagger}) F_{ij} c_i^{\dagger} c_j - \sum_i D_i(a_{\nu}; a_{\nu}^{\dagger}) \langle c_i^{\dagger} c_i \rangle, \end{aligned} \quad (\text{III.A.3.3})$$

where  $\sum_i D_i [c_i^{\dagger} c_i - \langle c_i^{\dagger} c_i \rangle] = U \mathbf{H}_{vb} U^{\dagger} - \mathbf{H}_{ph}$ . The effective matrix elements  $V_{ij}$  have the same form as the  $\bar{V}_{ij}$ , but with  $\bar{a}$  replaced by  $a$ . Throughout the rest of this work I shall make the Condon approximation, in which the  $V_{ij}$  are  $c$ -numbers,

although the methods which follow can be extended to non-Condon models. Similarly I shall restrict my attention to the harmonic Hamiltonians of Eq. (III.A.3.3), although anharmonicities can be included perturbatively.

The new elements in Eq. (III.A.3.3) are the  $F_{ij}$ , which are defined by  $F_{ij} = U c_i^\dagger c_j U^\dagger$ , so that, since  $U$  depends on  $c_i^\dagger c_i$ ,

$$F_{ij} = U(c_i^\dagger c_i \rightarrow 1; c_{k \neq i}^\dagger c_k \rightarrow 0) U^\dagger(c_j^\dagger c_j \rightarrow 1; c_{k \neq j}^\dagger c_k \rightarrow 0). \quad (\text{III.A.3.4})$$

The normally ordered form of  $F_{ij}$ , denoted by  $:F_{ij}:$ , may be found by evaluating the commutator of  $F$  with each of the  $a_\nu$  and solving the equations<sup>45</sup>

$$[:F_{ij}:, a_\nu] = -\frac{\partial :F_{ij}:}{\partial a_\nu^\dagger} \quad \text{and} \quad [:F_{ij}:, a_\nu^\dagger] = \frac{\partial :F_{ij}:}{\partial a_\nu}. \quad (\text{III.A.3.5})$$

Explicit results are

$$\begin{aligned} :F_{ij}: = K_{ij} \exp \left\{ \sum_{\mu\nu} [a_\mu^\dagger (R_{\mu\nu}^{-1}(ij) - \delta_{\mu\nu}) a_\nu + \frac{1}{2} (\sum_\lambda T_{\mu\lambda}(ij) R_{\lambda\nu}^{-1}(ij)) a_\nu \right. \\ \left. - \frac{1}{2} a_\mu^\dagger (\sum_\lambda R_{\mu\lambda}^{-1}(ij) T_{\lambda\nu}(ij)) a_\nu^\dagger - a_\mu^\dagger R_{\mu\nu}^{-1}(ij) \eta_\nu(j) \right. \\ \left. + a_\mu (\delta_{\mu\nu} - \sum_\lambda T_{\mu\lambda}(ij) R_{\lambda\nu}^{-1}(ij)) \eta_\nu(j) \right\}; \end{aligned} \quad (\text{III.A.3.6})$$

where the  $R_{\mu\nu}(ij)$ ,  $T_{\mu\nu}(ij)$ , and  $\eta_\mu(j)$  are related to the parameters of Eq's. (III.A.2.9) by

$$\begin{aligned} R_{\mu\nu}(ij) = \sum_{\lambda\delta\sigma} A_{\mu\lambda} (\langle c_k^\dagger c_k \rangle) A_{\lambda\sigma}^{-1} (c_i^\dagger c_i \rightarrow 1; c_{k \neq i}^\dagger c_k \rightarrow 0) \\ \times A_{\delta\sigma} (c_j^\dagger c_j \rightarrow 1; c_{k \neq j}^\dagger c_k \rightarrow 0) A_{\sigma\nu}^{-1} (\langle c_k^\dagger c_k \rangle) \\ \times \cosh \left\{ \ln \left[ \frac{\omega_\mu (\bar{\omega}_\mu - \rho_{\mu i} + \rho_{\mu j})}{\omega_\nu (\bar{\omega}_\nu + \rho_{\nu i} - \rho_{\nu j})} \right] \right\}; \end{aligned} \quad (\text{III.A.3.7})$$

$$T_{\mu\nu}(ij) = R_{\mu\nu}(ij; \cosh \rightarrow \sinh), \text{ and} \quad (\text{III.A.3.8})$$

$$\begin{aligned} \eta_{\mu}(ij) = \omega_{\mu}^{1/2} [ & \sum_{\nu} A_{\mu\nu} (\langle c_k^{\dagger} c_k \rangle) (\langle c_i^{\dagger} c_i \rangle - \langle c_j^{\dagger} c_j \rangle - 1) (\Delta_{\nu i} - \Delta_{\nu j}) \\ & - \sum_{\nu\lambda\sigma} A_{\mu\nu} (\langle c_k^{\dagger} c_k \rangle) A_{\nu\lambda}^{-1} (c_i^{\dagger} c_i \rightarrow 1; c_{k \neq i}^{\dagger} c_k \rightarrow 0) \\ & \times A_{\lambda\sigma} (c_j^{\dagger} c_j \rightarrow 1; c_{k \neq j}^{\dagger} c_k \rightarrow 0) \left[ \frac{\bar{\omega}_{\sigma} - \rho_{\sigma i} + \rho_{\sigma j}}{\bar{\omega}_{\sigma} + \rho_{\sigma i} - \rho_{\sigma j}} \right] \\ & \times (\langle c_i^{\dagger} c_i \rangle - \langle c_j^{\dagger} c_j \rangle + 1) (\Delta_{\sigma i} - \Delta_{\sigma j}) ]. \end{aligned} \quad (\text{III.A.3.9})$$

The constant  $K_{ij}$  may be determined from unitarity.

A considerable simplification of these results obtains if the atoms only undergo changes in equilibrium position as the electronic states change; that is if their normal mode structures and vibrational frequencies remain the same. In this case it may be shown that the  $D_i$  are zero, and the  $F_{ij}$  become

$$:F_{ij}: = K_{ij} \exp \left\{ \sum_{\nu} \eta_{\nu}(ij) (a_{\nu} - a_{\nu}^{\dagger}) \right\}; \quad (\text{III.A.3.10})$$

where  $K_{ij} = \exp[-1/2 \sum_{\nu} \eta_{\nu}^2(ij)]$  and

$$\eta_{\nu}(ij) = \omega_{\nu}^{1/2} \sum_{\mu} A_{\nu\mu} (\Delta_{\mu i} - \Delta_{\mu j}). \quad (\text{III.A.3.11})$$

It is clear that  $F_{ij}$  is a displacement operator that moves the coordinates of the atoms from one equilibrium position to the other when the electronic state changes from  $j$  to  $i$ . The general results of Eq. (III.A.3.6) not only account for this displacement but also provide a re-diagonalization of the normal modes and a rescaling of the coordinates to take account of the frequency changes.

Summarizing all of these results, the general Hamiltonian can be written as

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_{int} \quad (\text{III.A.3.12})$$

where the zero-order Hamiltonian consists of decoupled electron and phonon degrees of freedom, while the interaction Hamiltonian has the general form

$$\mathbf{H}_{int} = \sum_{ij} M_{ij} c_i^\dagger c_j, \quad (\text{III.A.3.13})$$

where the  $M_{ij}$  are functions of the phonon operators. In the special case where there are no frequency shifts or normal mode re-diagonalizations, the  $M_{ij}$  are zero if  $i = j$ ; there are no "diagonal" terms in the interaction. In the following paragraphs I present a general method for calculating the electron or phonon dynamics in systems described by this form of electron-phonon coupling. Before attempting the general treatment, it is useful to solve the simplest possible case, namely two electronic states coupled to a single phonon mode.

A model Hamiltonian with two electronic states and a single phonon mode can be represented by the adiabatic potential surfaces of Fig. III-8. For every vibrational state of one electronic level there is a corresponding, nearly degenerate vibrational level of the other electronic state; all other levels are separated by at least  $\hbar\omega$ , where  $\omega$  is the mode frequency. Unless the matrix element  $V$  is comparable to this separation, only the nearly degenerate levels are coupled by the interaction Hamiltonian. In the absence of vibrational relaxation, the system collapses to an infinite set of two level systems, which can be solved trivially. In particular, at zero temperature, if the system starts in the upper electronic state, only one of these two level systems is accessible, and the population simply oscillates back and forth between the two levels of interest. Dissipation may be included by noting, again at zero temperature, that vibrational relaxation serves only to irreversibly deplete the population of one of the states, as schematized in Fig. III-8.

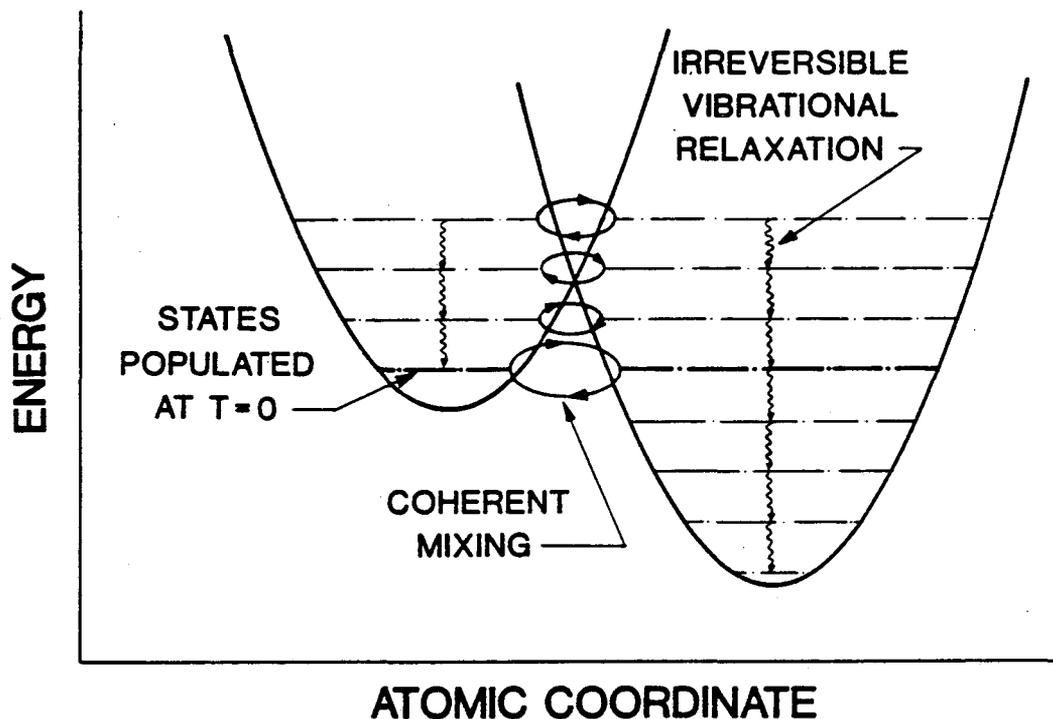


FIGURE III-8 Electronic transitions and vibrational relaxation. The simplest example of the model Hamiltonians considered in the text, with two electronic states and one phonon mode which is coupled only *via* a displacement shift between the two states, as shown. Each vibrational level is coupled *coherently* to its quasi-degenerate partner by the electronic matrix element, with the coupling being modified by a Franck-Condon overlap factor. Vibrational relaxation, which is an effectively irreversible coupling of the vibrational modes to a heat bath, induces incoherent transitions among the vibrational levels of each electronic state. This picture, which emphasizes the interplay between coherent and dissipative processes should be contrasted with that of Fig. III-4.

Figure III-8 suggests that the amplitudes  $\phi_1$  and  $\phi_2$  for the system to be in the two states obey the equations

$$i\dot{\phi}_1 = \epsilon\phi_1 + V\phi_2 \tag{III.A.3.14a}$$

$$i\dot{\phi}_2 = V\phi_1 + (\epsilon - i\Gamma)\phi_2, \tag{III.A.3.14b}$$

where  $V$  is a matrix element and  $\Gamma$  is the vibrational relaxation rate which depletes the final state; I assume, for simplicity, exact degeneracy of the two states. The dynamics of the electronic states can be found from Green's functions such as (see below)

$$G_{eff}(\Omega) = \int dt e^{i\Omega t} \langle \phi_1(t) \phi_1^\dagger(0) \rangle \theta(t) \quad (\text{III.A.3.15})$$

where  $\theta(t)$  is the unit step function. By solving the initial value problem for Eq's. (III.A.3.14), it may be seen that

$$G_{eff}(\Omega) \approx \frac{1}{\Omega_+ - \Omega_-} \left\{ \frac{\Omega_+}{\Omega - \Omega_-} - \frac{\Omega_-}{\Omega - \Omega_+} \right\}, \quad (\text{III.A.3.16})$$

where the eigenfrequencies, corresponding to poles of the Green's function, are given by

$$\Omega_{\pm} = \epsilon + \frac{1}{2} [i\Gamma \pm \sqrt{4V^2 - \Gamma^2}]. \quad (\text{III.A.3.17})$$

If the relaxation rate is small,  $\Gamma \ll V$ , this reduces to  $\Omega_{\pm} = \epsilon \pm V$ , so that the poles remain at real frequencies. As a result, the dynamics are purely oscillatory, or coherent; if we start the system in one state it oscillates back and forth to the other accessible state, as concluded above from an intuitive argument. On the other hand, if the relaxation rate is large, both poles are shifted off the real axis, one by the relaxation rate itself, and one by the apparent decay rate  $T_2^{-1} = V^2/\Gamma$ . The pole at the relaxation rate, however, has vanishing residue in this limit, as may be shown directly from Eq. (III.A.3.16), and therefore does not contribute to the electronic Green's function--in the slow regime, there are *no* dynamics of the electronic states on the time scale of vibrational relaxation. Note that the decay rate in the slow regime is just what we would obtain by applying the golden rule to the transition, as it must be. Thus we see that, as the relaxation rate varies, we expect a cross-over from purely real self-energies and oscillatory behavior of the populations to pure imaginary self-energies and simple exponential decays whose rates are consistent with those calculated by naive perturbation theory. With these results in mind, let us turn to a more general approach.

As is well known, all of the experimental observables in the system can be related to Green's functions of the various operators in the model Hamiltonian.<sup>48</sup> Thus, if we wish to know the lifetime of the state created by  $c_i^\dagger$ , then we can calculate the Green's function<sup>49</sup>

$$G_{ii}(\Omega) = -i \int d\tau e^{i\Omega\tau} \langle \mathbf{T}[\bar{c}_i(t)\bar{c}_i^\dagger(t-\tau)] \rangle, \quad (\text{III.A.3.18})$$

where  $\bar{c}$  denotes a Heisenberg picture operator, and examine its poles in the complex  $\Omega$  plane; the real part of the pole frequency is the energy of the excitation and the imaginary part is the decay rate.<sup>50</sup>

Perturbation theory for the Green's functions is based on the fact that the averages of Heisenberg picture operators can be written in terms of interaction picture operators whose properties are calculated from the zero-order Hamiltonian. For example,

$$G_{ii}(\Omega) = -i \int d\tau e^{i\Omega\tau} \frac{\langle \mathbf{T}[S c_i^\dagger(t) c_i(t+\tau)] \rangle}{\langle \mathbf{T}[S] \rangle}, \quad (\text{III.A.3.19})$$

$$S = \exp\left\{-i \int dt \mathbf{H}_{int}(t)\right\}, \quad (\text{III.A.3.20})$$

where in these equations, and throughout the remainder of this section, the time dependent operators  $c, \mathbf{H}_{int}$ , etc., are in the interaction picture.

The problem of calculating these Green's functions can be considerably simplified by a series of transformations which relate the quantities of interest to integrals over a generating functional. In this way we can make use of the functional integral methods of field theory<sup>51-53</sup> to derive both exact and perturbative results. To begin, we note that

$$G_{ij}(\Omega) = -i \int d\tau e^{+i\Omega\tau} \frac{\delta^2 \Lambda_G[J]}{\delta J_i(t-\tau) \delta J_j^\dagger(t)} \Big|_{J=0}, \quad (\text{III.A.3.21})$$

where  $\delta/\delta J_i(t)$  denotes functional differentiation, and the generating functional  $\Lambda_G$  is

$$\Lambda_G[J] = \frac{\langle T[\exp\{-i \int dt [\mathbf{H}_{int}(t) + \sum_i (J_i^\dagger(t) c_i(t) + J_i(t) c_i^\dagger(t))\}] \rangle}{\langle T[\exp\{-i \int dt [\mathbf{H}_{int}(t)] \rangle} \quad (III.A.3.22)$$

This equation is not correct as it stands, since it will not reproduce the relations among higher-order Green's functions which are required by the anti-commutativity of  $c_i$  and  $c_i^\dagger$ . This problem can be solved by taking the variables  $J$  and  $J^\dagger$  themselves to anti-commute, or be elements of a Grassman algebra.<sup>52,53</sup> This technique is well known in field theory, and is useful primarily because integration over a Gaussian function of Grassman variables is almost identical to the corresponding integral over real variables:

$$\int d\eta^\dagger d\eta \exp[i\eta^\dagger M \eta + i\zeta^\dagger \eta + i\zeta \eta^\dagger] = (\det M) e^{-i\zeta^\dagger M^{-1} \zeta},$$

where  $\eta$  and  $\zeta$  are vectors of complex Grassman numbers,  $\eta^\dagger$  and  $\zeta^\dagger$  are their Hermitian conjugates, and  $M$  is a matrix of complex (ordinary) numbers. With this result, we can proceed without further difficulty. In particular, note that in Eq. (III.A.3.22) the denominator serves only to normalize the generating functional to  $\Lambda[J \rightarrow 0] = 1$ , and may be discarded if this normalization is restored at the end of the calculation; I shall refer to the un-normalized functional as  $\Lambda'$ .

The fact that we can integrate over Grassman variables means that we can apply some basic results from Fourier analysis. In particular, if have some functional  $\chi$  of the operators  $c_i$  and  $c_i^\dagger$ , then

$$\langle T(\chi[c_i, c_i^\dagger]) \rangle = \int [d\mu] \int [d\nu] e^{i\mu \cdot \nu} \langle T(e^{-i\nu \cdot c}) \rangle \chi[\mu_i, \mu_i^\dagger], \quad (III.A.3.23)$$

where I have used the shorthand notation

$$\mu \cdot \nu = \int dt \sum_i [\mu_i^\dagger(t) \nu_i(t) + \mu_i(t) \nu_i^\dagger(t)], \quad (III.A.3.24)$$

and  $\int [d\mu]$  denotes functional integration over the variable  $\mu$ . Thus the generating functional becomes

$$\begin{aligned} \Lambda'[J] &= \int [d\mu] \int [d\nu] e^{i\mu \cdot \nu} e^{iJ \cdot \mu} \langle T(e^{-i\nu \cdot c}) \rangle \\ &\times \langle T(\exp\{-i \int dt H_{int}[\mu_i(t), \mu_j^\dagger(t)]\}) \rangle, \end{aligned} \quad (\text{III.A.3.25})$$

where the phonon operators are still buried in the definition of  $H_{int}$

The functional  $\langle T(e^{-i\nu \cdot c}) \rangle$  is clearly the same as  $\Lambda_G$ , but with  $H_{int} \rightarrow 0$ . Thus this functional generates the electronic Green's functions in the absence of the interaction Hamiltonian, or the zero-order Green's functions. By Wick's theorem the zero-order Green's functions always factor into products of the single particle Green's function  $G^{(0)}$ , so that their generating functional is Gaussian. Thus we must have

$$\langle T(e^{-i\nu \cdot c}) \rangle = e^{i\nu^\dagger G^{(0)} \cdot \nu}, \quad (\text{III.A.3.26})$$

where, by analogy with the notation introduced in Eq. (III.A.3.24),

$$\nu^\dagger \cdot G^{(0)} \cdot \nu = \int dt \int dt' \sum_{ij} \nu_i^\dagger(t) G_{ij}^{(0)}(t'-t) \nu_j(t). \quad (\text{III.A.3.27})$$

Returning to Eq. (III.A.3.25), it is clear that the physics of the model Hamiltonian is contained in the functional

$$\Xi[\mu] = \langle T(\exp\{-i \int dt H_{int}[\mu_i(t), \mu_j^\dagger(t)]\}) \rangle. \quad (\text{III.A.3.28})$$

For the Hamiltonians of Eq. (III.A.3.13), this becomes

$$\Xi[\mu] = \langle T(\exp\{-i \int dt \mu_j^\dagger(t) M_{ij}(t) \mu_i(t)\}) \rangle, \quad (\text{III.A.3.29})$$

where I adopt the summation convention for repeated indices  $i$  and  $j$ . The average over the phonon operators which remain in the  $M_{ij}$  can be done by the cumulant expansion,<sup>54</sup> which will give us an exponential of a power series in  $V$ , where  $V$

measures a typical electronic matrix element in  $H_{int}$ . To lowest non-vanishing order<sup>55</sup> one obtains

$$\Xi^{(1)}[\mu] = \exp\left\{-\frac{1}{2} \sum_{i,j,k,l} \int dt \int dt' \langle T[M_{ij}(t)M_{kl}(t')] \rangle \right. \\ \left. \times \mu_i^\dagger(t)\mu_j(t)\mu_k^\dagger(t')\mu_l(t')\right\}. \quad (\text{III.A.3.30})$$

The generating functional which we find upon combining all these results is clearly the same as would be obtained if the interaction Hamiltonian involved no phonons and were defined by

$$\int dt H_{eff}(t) = \frac{i}{2} \int dt \int dt' \sum_{i,j,k,l} \langle T[M_{ij}(t)M_{kl}(t')] \rangle \\ \times c_i^\dagger(t)c_j(t)c_k^\dagger(t')c_l(t'). \quad (\text{III.A.3.31})$$

This Hamiltonian corresponds to an effective pairwise interaction among the quasi-particles created by the  $c_i^\dagger$ . In the usual pairwise interaction problems of many-body theory,<sup>48,56</sup> the indices denote the spin states, and there is a spatial dependence which does not arise in the molecular problems treated here. The lack of spatial dependence means that the usual four-momentum integrals collapse to frequency integrals, but the (nominal) spin sums become non-trivial. With these differences in mind, the “momentum space” Feynman rules for the effective Hamiltonian of Eq. (III.A.3.31) are:

- [1] Draw all topologically distinct diagrams with  $n$  interaction lines, and hence  $2n$  vertices; these diagrams represent the  $n^{\text{th}}$  order of perturbation theory.
- [2] Assign a directed frequency to both interaction and Green's function lines, and conserve frequency at each vertex.

[3] For each Green's function line, insert

$$G_{ij}^{(0)}(\Omega) = \delta_{ij} \left\{ \frac{1 - n_i}{\Omega - \epsilon_i + i\delta} + \frac{n_i}{\Omega - \epsilon_i - i\delta} \right\}, \quad (\text{III.A.3.32})$$

where  $\epsilon_i$  and  $n_i$  are the zero-order energy and ground-state occupancy of the state  $i$ , respectively, while  $\delta \rightarrow 0^+$  at the end of the calculation.

Note that the  $n_i$  are not completely determined by the  $\epsilon_i$  since we have not specified the chemical potential of the system; this may vary depending on what we are trying to calculate. Thus to examine the decay rates out of the various electronic states, it is convenient (though not necessary) to assume that none of the states are populated and hence  $n_i = 0$  for all  $i$ . On the other hand, this is inappropriate if we are interested in the effects of the electronic coupling on the phonon Green's functions, or in the radiative interactions of the system, and in these cases we must use the more physical choice corresponding to unit population of the ground state.

[4] For each interaction line insert

$$D_{ijkl}(\Omega) = -i \int d\tau e^{i\Omega\tau} \langle \text{T}[M_{ij}(t+\tau)M_{kl}(t)] \rangle. \quad (\text{III.A.3.33})$$

[5] Sum over all free indices; integrate over all free frequencies. Integration of a

function  $Z(\Omega)$  over  $\Omega$  is understood to mean  $\int \frac{d\Omega}{2\pi} Z(\Omega)$ .

[6] Multiply by a factor  $i^n (-1)^f$ , where  $f$  is the number of closed Green's function loops; these loops are to be interpreted as  $e^{i\Omega\delta} G_{ij}(\Omega)$ , with  $\delta \rightarrow 0^+$ .

To calculate the single-particle Green's functions we make use of Dyson's equation,

$$G_{ij}(\Omega) = G_{ij}^{(0)}(\Omega) + G_{il}^{(0)}(\Omega) \Sigma_{lk}(\Omega) G_{kj}(\Omega), \quad (\text{III.A.3.34})$$

where the lowest order contributions to the proper self energy  $\Sigma_{lk}$  are shown in Fig.

III-9. These diagrams correspond to the analytic forms

$$\Sigma_{lk}^{(1a)}(\Omega) = -i \int \frac{d\Omega'}{2\pi} D_{lk\gamma\delta}(0) G_{\gamma\delta}^{(0)}(\Omega') e^{i\Omega'\eta}, \text{ and} \quad (\text{III.A.3.35a})$$

$$\Sigma_{lk}^{(1b)}(\Omega) = i \int \frac{d\Omega'}{2\pi} D_{l\gamma\delta k}(\Omega') G_{\gamma\delta}^{(0)}(\Omega - \Omega'). \quad (\text{III.A.3.35b})$$

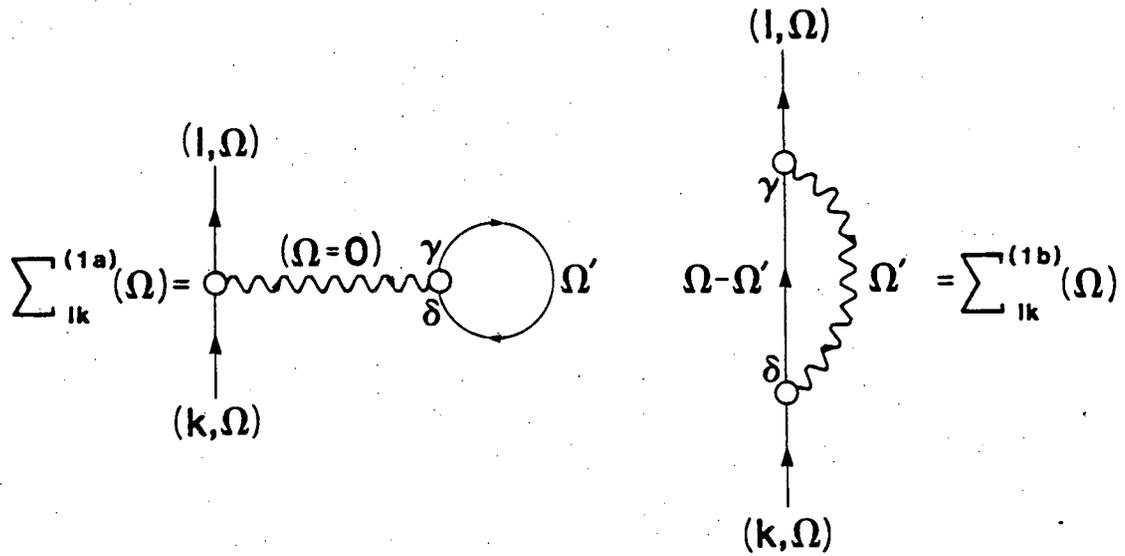


FIGURE III-9 Lowest-order contributions to the electronic self-energy, Eq's. (III.A.3.35).

Similarly, the two-particle Green's function

$$G_{ijkl}(t_1, t_2; t_3, t_4) = \frac{1}{\langle \mathbf{T}[S] \rangle} \langle \mathbf{T}[c_i(t_1) c_j(t_2) c_k^\dagger(t_3) c_l^\dagger(t_4) S] \rangle \quad (\text{III.A.3.36})$$

is given by<sup>56</sup>

$$\begin{aligned} G_{ijkl}(t_1, t_2; t_3, t_4) &= G_{ki}(t_3 - t_1) G_{lj}(t_4 - t_2) - G_{kj}(t_3 - t_2) G_{li}(t_4 - t_1) \\ &+ i \int d^4\tau G_{\gamma_1 i}(\tau_1 - t_1) G_{\gamma_2 j}(\tau_2 - t_2) G_{\gamma_3 k}(\tau_3 - t_3) G_{\gamma_4 l}(\tau_4 - t_4) \\ &\times \Gamma_{\gamma_1 \gamma_2; \gamma_3 \gamma_4}(\tau_1, \tau_2; \tau_3, \tau_4), \end{aligned} \quad (\text{III.A.3.37})$$

where  $\Gamma$  is the vertex part defined according to Fig. III-10.

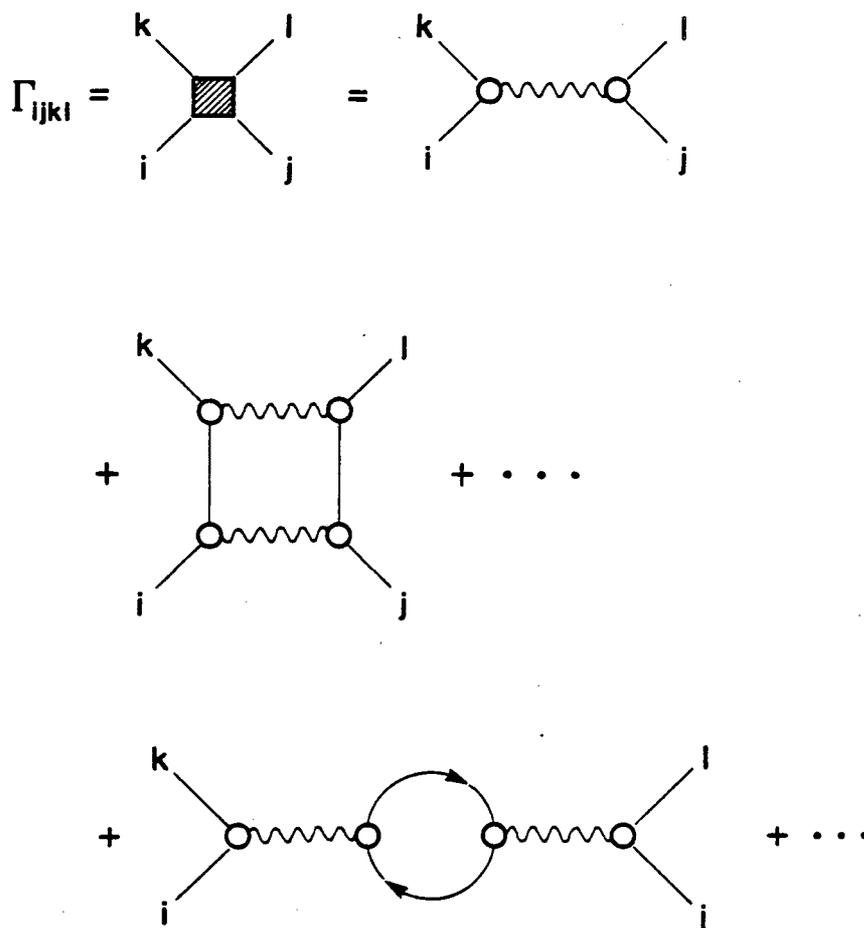


FIGURE III-10 Lowest-order contributions to the electronic vertex part. Top: Zero-order term. Middle: First ladder correction. Bottom: First ring correction.

In order to use these results we must have some way of computing a time-ordered expectation value of the phonon operators such as  $\langle T[M_{ij}(t)M_{kl}(t')] \rangle$ . By analogy with what has been done to the electronic Green's functions we may make use of a generating function technique. The result is that

$$\begin{aligned}
 & \langle T[:F_1(a_1^\dagger, a_1)::F_2(a_2^\dagger, a_2)::] \rangle \\
 &= \int d^4\alpha d^4\xi e^{-i\xi\bar{\alpha}} \langle T[e^{-i\xi_1^\dagger a_1^\dagger} e^{-i\xi_1 a_1} e^{-i\xi_2^\dagger a_2^\dagger} e^{-i\xi_2 a_2}] \rangle \\
 & \quad \times :F_1(\alpha_1^\dagger, \alpha_1)::F_2(\alpha_2^\dagger, \alpha_2)::,
 \end{aligned}
 \tag{III.A.3.38}$$

where  $:F_1:$  and  $:F_2:$  are normally ordered functions of the phonon creation and annihilation operators  $a_1 = a(t_1)$ , etc., the  $\bar{\xi}$  and  $\bar{\alpha}$  are vectors of  $c$ -numbers, and the generating function is given by

$$\begin{aligned} & \langle \mathbf{T}[e^{-i\xi_1^\dagger a_1^\dagger} e^{-i\xi_1 a_1} e^{-i\xi_2^\dagger a_2^\dagger} e^{-i\xi_2 a_2}] \rangle \\ & = \exp\left\{-\left(\xi_1^\dagger \xi_2 + \xi_2^\dagger \xi_1\right) e^{-(i\omega - \gamma)|\tau|}\right\}, \end{aligned} \quad (\text{III.A.3.39})$$

where  $\tau = t_1 - t_2$ .

These techniques can be illustrated by the simplest example of this family of Hamiltonians, consisting again of two electronic states coupled to a single phonon mode with no frequency shift. We are interested in the self-energy of the upper level, since at zero-temperature the ground state is infinitely long-lived. The interaction Hamiltonian in this case is simply

$$\mathbf{H}_{int} = VF^\dagger c_2^\dagger c_1 + VFc_1^\dagger c_2, \quad (\text{III.A.3.40})$$

where  $c_2^\dagger$  creates molecule in the upper level,  $c_1^\dagger$  creates molecules in the lower level, and from Eq. (III.A.3.10) the operator  $F$  is

$$F = e^{-1/2\eta^2} e^{\eta(a - a^\dagger)}. \quad (\text{III.A.3.41})$$

It is clear that the only non-vanishing interaction terms are

$$D_{1212}(\Omega) = -iV^2 \int d\tau e^{i\Omega\tau} \langle \mathbf{T}[F(\tau)F(0)] \rangle, \quad (\text{III.A.3.42a})$$

$$D_{2112}(\Omega) = -iV^2 \int d\tau e^{i\Omega\tau} \langle \mathbf{T}[F^\dagger(\tau)F(0)] \rangle, \quad (\text{III.A.3.42b})$$

$$D_{2121}(\Omega) = -iV^2 \int d\tau e^{i\Omega\tau} \langle \mathbf{T}[F^\dagger(\tau)F^\dagger(0)] \rangle, \text{ and} \quad (\text{III.A.3.42c})$$

$$D_{1221}(\Omega) = -iV^2 \int d\tau e^{i\Omega\tau} \langle \mathbf{T}[F(\tau)F^\dagger(0)] \rangle, \quad (\text{III.A.3.42d})$$

so that the self-energy term  $\Sigma_{22}^{(1a)} \approx D_{22\gamma\delta}$  is identically zero. The remaining term is

$$\Sigma_{22}^{(1b)}(\Omega) = i \int \frac{d\Omega'}{2\pi} D_{2112}(\Omega') G_{11}^{(0)}(\Omega - \Omega'). \quad (\text{III.A.3.43})$$

When all the dust clears, this becomes

$$\Sigma_{22}^{(1b)}(\Omega) = V^2 e^{-S} \sum_{n=0}^{\infty} \frac{S^n}{n!} \frac{1}{(\Omega - \epsilon_1 - n\omega) - i\gamma n}, \quad (\text{III.A.3.44})$$

where  $S = \eta^2$ , and  $\epsilon_1$  is the energy of the lower state. By solving for the poles of the Green's function in the presence of this self-energy, it is possible to trace the "trajectories" of these poles as the ratio  $V/\gamma$  changes; an example of the behavior obtained in this way is shown in Fig. III-11. It is clear that a cross-over from oscillatory to exponentially damped behavior is observed, as predicted previously, and it is possible to delineate from such plots the boundaries of the two regimes.

The next point about the simplest model concerns the multi-particle Green's functions. In particular, the two particle Green's function with all indices equal describes the time course of the number of molecules in a given electronic state,  $G_{iiii}(\tau) \approx \langle n_i(\tau) n_i(0) \rangle$ . This Green's function is given by the sum of its reducible part, which is just a product of one-particle Green's functions, and its irreducible part, which is related to the vertex part as in Eq. (III.A.3.37). It is easy to show, however, that in this model the irreducible part is of order  $V^2$ . This implies that, to lowest order in perturbation theory, the decay of the number operator occurs exactly twice as fast as the decay of the creation operator. Since the expectation value of the creation operator is a measure of the coherence between the upper and lower states, this implies that coherence decays at a rate one-half that describing the decay of the population in the upper state. To understand the significance of this result, we must look at some of the basic features of two-level systems.

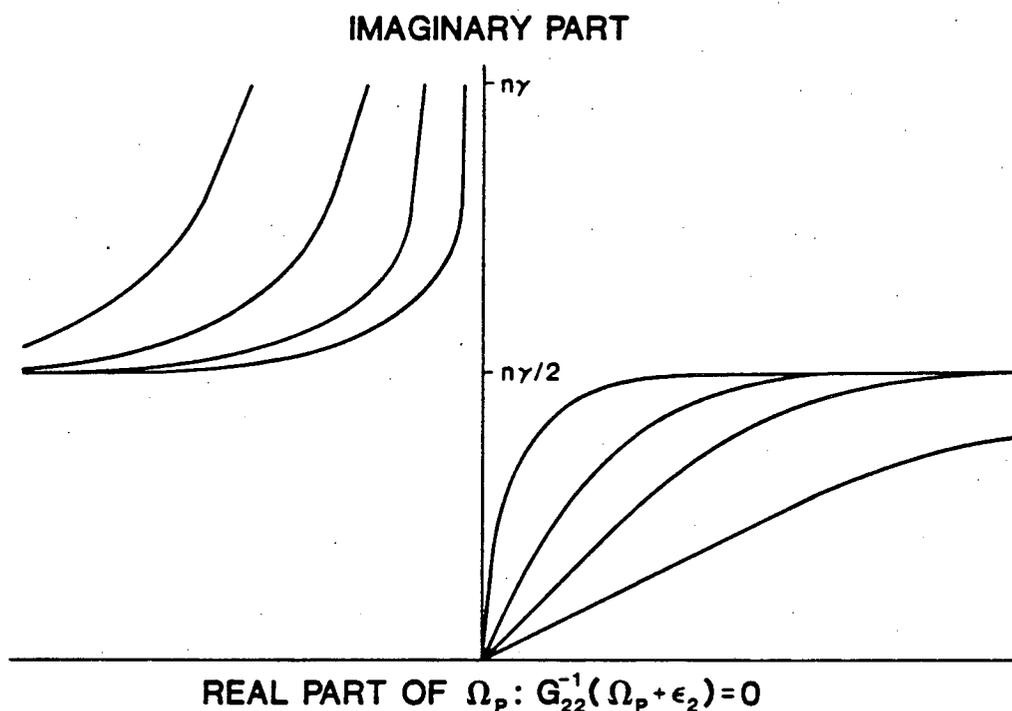


FIGURE III-11 Coherent vs. incoherent behavior in a Green's function analysis of electronic transitions. "Nyquist plots" of the real vs. imaginary parts of the pole frequencies for the Green's function  $G_{22}$ , with the electronic matrix element as a parameter which increases in the direction of the arrows. For small values of the matrix element, the smaller pole frequency is pure imaginary and corresponds to the decay rate calculated from simple perturbation theory, while as the matrix element increases this pole moves off the imaginary axis, reflecting partially coherent behavior; different curves correspond to different values of the "detuning parameter"  $\epsilon_2 - \epsilon_1 - n\omega$ , with the restriction that this parameter is positive, which simplifies the diagram. The larger pole also moves from being approximately imaginary and equal to the vibrational relaxation rate to being largely real, reflecting coherent evolution of the electronic states.

The theory of relaxation in two-level systems<sup>43,57</sup> is based on a Hamiltonian of the form

$$\mathbf{H} = \frac{\epsilon'}{2}\sigma_z + H_x(t)\sigma_x + H_y(t)\sigma_y + H_z(t)\sigma_z, \quad (\text{III.A.3.45})$$

where the  $\sigma_i$  are Pauli matrices such that the eigenstates of  $\sigma_z$  correspond to states 1 and 2 defined above. The  $H_i(t)$  are stochastic functions of time, representing some random magnetic field to which the spin is subject, and may be interpreted more

rigorously as analogous to the heat bath coordinate  $\Gamma$  in Eq. (III.A.2.3). If the heat bath remains in equilibrium throughout the change in electronic state, and if measurements are made on a time scale long compared to the correlation times of the  $H_i(t)$ , then the expectation values of the  $\sigma$  operators obey the Bloch equations:

$$\langle \dot{\sigma}_x \rangle = -\frac{\langle \sigma_x \rangle}{T_2} + \epsilon' \langle \sigma_y \rangle, \quad (\text{III.A.3.46a})$$

$$\langle \dot{\sigma}_y \rangle = -\frac{\langle \sigma_y \rangle}{T_2} - \epsilon' \langle \sigma_x \rangle, \text{ and} \quad (\text{III.A.3.46b})$$

$$\langle \dot{\sigma}_z \rangle = -\frac{\langle \sigma_z \rangle - \sigma_z(\infty)}{T_1}. \quad (\text{III.A.3.46c})$$

To lowest order in perturbation theory,

$$T_1^{-1} = \frac{2}{\hbar^2} [k_{xx}(\epsilon') + k_{yy}(\epsilon')], \quad (\text{III.A.3.47a})$$

$$T_2^{-1} = \frac{1}{\hbar^2} [2k_{zz}(0) + k_{xx}(\epsilon') + k_{yy}(\epsilon')], \quad (\text{III.A.3.47b})$$

$$k_{ii}(\epsilon) = \int_{-\infty}^{\infty} d\tau \langle H_i(t) H_i(t+\tau) \rangle e^{-\frac{\epsilon}{\hbar} \tau}. \quad (\text{III.A.3.47c})$$

Note that  $T_1$  is the time constant for relaxation of the electronic populations, while  $T_2$  is the time constant for the relaxation of the coherence between electronic states which is necessary to generate non-zero average "magnetization" in the  $x$ - $y$  plane.

From the Bloch equations it is apparent that the results derived above imply a relation between  $T_2$  and  $T_1$ , namely  $T_2 = 2T_1$ . This means that in our simple model, the only contributions to dephasing arise from lifetime broadening--there are no pure dephasing terms. But this conclusion, which is obtained here by field-theoretic methods and is independent of the time scale of the transitions among electronic

states, is exactly the same as obtained from the Bloch equations when there are no fluctuating fields in the  $z$  direction, which corresponds to terms diagonal in electron operators. Indeed there are no diagonal terms in the model Hamiltonian of Eq. (III.A.3.15), so that the Green's function method generalizes the results of the "slow regime," where the Bloch equations are valid, to all time regimes.

In perturbation theory at least, the identification of pure dephasing with diagonal terms in the interaction Hamiltonian is independent of the time scale of the radiationless dynamics. Correspondingly perturbation theory predicts that coherence must last for the full time course of the electronic transition in the absence of such terms, or if these terms are small, as will be the case at low temperature. This correspondence of results from the Bloch equations and field-theoretic methods suggests that the dynamics of radiationless transitions may be calculated directly from relaxation theory, subject to the constraint that the dynamics be slow compared to vibrational relaxation. Such calculations<sup>58</sup> are described in Appendix D.

Some more general results about the relations between dephasing and population decay can be obtained by interpreting the terms of the diagrammatic perturbation series. To lowest order we are only interested in the diagonal elements of the self-energy matrix, since the zero-order Green's functions are themselves diagonal. In this case the contribution from Eq. (III.A.3.35a) to the self-energy of state  $j$  is  $\approx \sum_k \langle M_{jj} M_{kk} \rangle$ , which is a measure of the fluctuations in energy of state  $j$  relative to the other states of the system. By analogy with the discussion of the Bloch equations, this is a pure dephasing term: it causes decay of the single-particle Green's function but does not arise from matrix elements which can cause transitions among states. The contribution of these terms to the dephasing rate is proportional to their spectral density at zero frequency, as was true also in the Bloch equation treatment.

The term in Eq. (III.A.3.35b) is of the form  $\sum_k \langle M_{jk} M_{kj} \rangle$ , and thus represents the effects of transitions between states  $j$  and  $k$ . The naive perturbation

theory result for the transition rate between these two states would be  $\approx \sum_k D_{jkkj}(\epsilon_j - \epsilon_k)$ , and this would also be the result of the Bloch equation treatment [cf. Eq. (III.A.3.47)]. The results of Eq's. (III.A.3.35) generalize the naive perturbation theory to include processes which occur "off the mass shell," since the energies are not fixed to their zero-order values. To understand why this is the appropriate generalization, we must consider carefully the assumptions built into the Bloch equations.

The Bloch equations are valid only on time scales long compared to the correlation times for the fluctuating fields. If we make observations on such long time scales  $T$ , the energy uncertainties are very small, of order  $\hbar/T$ . If we start to speed up the observations, energy uncertainties increase, and we are no longer guaranteed that the energies of the intermediate phonons or electrons are equal to their zero-order energies. As a result, the off-shell contributions become significant and, as in the oscillatory behavior derived above, can eventually dominate.

Combining the two terms of Fig. III-9, we find that the single-particle Green's function decays at a rate related to both the natural lifetimes of the states and to the pure dephasing terms. The two-particle Green's function, for appropriate values of the indices, corresponds strictly to the decay of the number operators for the states, and should therefore have no contribution from the dephasing terms. If there were no vertex part the one-particle and two-particle Green's functions would be simply related and the populations would decay at rate dependent upon dephasing processes, which is of course incorrect. In fact, if there are significant diagonal terms in the interaction Hamiltonian which generate the dephasing, then the vertex term of Eq. (III.A.3.37) which causes the one- and two-particle Green's functions to differ also become significant. It is thus through the vertex part that dephasing and lifetime broadening are disentangled.

At first glance it does not appear that the disentangling will work. The vertex term in Eq. (III.A.3.37) is additive, so that the two-particle Green's function consists of its reducible part plus the vertex term; it is precisely this structure which enabled us to conclude above that the irreducible part is of order  $V^2$  in the simplest model. At best it thus would seem that we obtain two terms which will decay at different rates. The answer to this problem can be seen by examining a simple case in which all the  $D_{ijkl}$  are frequency independent. This means that the  $M_{ij}$  have very short correlation times, and should therefore correspond well with the results of the Bloch equations.

The lowest order vertex term in the two particle Green's function with all indices equal becomes

$$\begin{aligned} \Delta G_{iiii}(t_1, t_2; t_3, t_4) = & iD_{iiii} \int d^4\tau G_{ii}^{(0)}(\tau_1 - t_1) G_{ii}^{(0)}(\tau_2 - t_2) \\ & \times G_{ii}^{(0)}(\tau_3 - t_3) G_{ii}^{(0)}(\tau_4 - t_4). \end{aligned} \quad (\text{III.A.3.48})$$

For particular choices of the  $t_n$ , this integral diverges. In fact it is but one of a family of terms, shown in Fig. III-10, which diverge; these are the ladder diagrams. It is only when these diagrams are summed that we can get a sensible answer, and then the vertex terms become of order unity and cancel the reducible parts just as they must.

The need to sum ladder diagrams is familiar from both many-body and particle theory. In the many-body problem, they are significant whenever the interactions are of short range but potentially strong, as is the case in the nuclear matter and other "hard-core" systems.<sup>59</sup> The analogy in the present problem is to the short correlation time of the fluctuating fields. In particle theory, the ladder diagrams are important in determining the low-energy behavior of the effective interactions (cf. Ref. 52). The analogy now is that low energies correspond to long observation intervals, and "long" is relative to the time scales intrinsic to the problem, namely the correlation times.

One could imagine doing calculations in the opposite regime, for time scales short compared to the correlation times of the fluctuating fields. In this case the analogies with many-body theory suggest that the effective interaction is long-ranged, so that the ring diagrams (cf. Fig. III-10) are the most strongly divergent.<sup>60</sup> Indeed one may verify this using the Feynman rules given above but with  $D_{ijkl}(\Omega) \rightarrow \delta(\Omega)$  rather than a constant. Again a proper disentangling of the dephasing and lifetime broadening requires summation of a class of terms to infinite order.

What we have found is that there is no consistent first order treatment of pure dephasing terms in the multi-particle Green's functions. This is true in spite of the fact that the contributions to the single-particle self-energies are just the natural generalizations of the Bloch equation results, which are a first order theory. This apparent contradiction may be traced to the fact that the generating functionals for the electronic Green's functions are non-Gaussian, with the size of the non-Gaussian terms being related to the magnitude of the pure dephasing.<sup>61</sup>

The non-Gaussian character of the generating functionals means that there is a non-linearity of the the response to external perturbations. In particular, if we imagine using an external field to drive the system to a far from equilibrium state, the time course of relaxation to equilibrium will depend on the strength of external field, or equivalently on the state of the system. In contrast, the Bloch equations are linear, and hence predict a relaxation process which is independent of the initial state. Thus we must conclude that the Bloch equations are not rigorously applicable to systems with pure dephasing, that is with  $T_2 \neq 2T_1$ . This conclusion is not restricted to the particular form of model Hamiltonian used in this work, and the inconsistencies between linear relaxation theory and the presence of pure dephasing may also be shown without the use of the Green's function techniques.<sup>62</sup>

Aside from providing a rigorous (if surprising) approach to the problems of dephasing, the field theoretical methods used here provide a method for calculating

any quantity of interest to any desired order, subject only to the initial approximation of perturbation theory in the electronic matrix element; this assumption can itself be relaxed with some effort. The calculations at zero temperature can be readily generalized by the use of temperature Green's functions, and it is possible in some models to give comparably simple treatments of systems in non-equilibrium states. If we wish to analyze spectroscopic experiments we must add to the model Hamiltonian appropriate terms for the radiation field and its interaction with the system, but the basic generating functional techniques still allow the derivation of an effective diagrammatic perturbation theory, as described in Appendix E. These methods allow us to extract several properties of the model Hamiltonians which have been overlooked or not treated rigorously by previous methods. The following section focuses on two such features of the theories, with an eye toward their biological significance.

#### 4. *Some predicted features of biomolecular dynamics*

Two important features emerge from the calculations presented in the previous section. First, as may be seen from Eq. (III.A.3.44), the self-energy of an electronic state--and hence the rates of transitions between electronic states--exhibits a resonant dependence on the vibrational frequencies. Second, as illustrated in Fig. II-11, for reactions which occur on a time scale comparable to the vibrational relaxation time the electronic states retain their quantum mechanical coherence.

The resonance effect is quite easy to understand, and can be generalized at least qualitatively to systems with multiple vibrational modes. If we have an electronic transition which must give up an energy  $\epsilon$  then the probability of this transition occurring with the emission of  $n_\mu$  phonons at each frequency  $\omega_\mu$  is proportional to

$$W \approx \frac{\sum_{\mu} \gamma_{\mu}}{(\epsilon - \sum_{\mu} n_{\mu} \omega_{\mu})^2 + (\sum_{\mu} n_{\mu} \gamma_{\mu})^2}, \quad (\text{III.A.4.1})$$

where  $\gamma_{\mu}$  is the natural linewidth of the  $\mu^{\text{th}}$  mode; this result is essentially just the

density of states factor in the Golden rule expression<sup>5</sup> for the transition rate. The rate is maximum at those energy gaps which can be written as integer combinations of the vibrational frequencies, which is a strictly quantum mechanical effect--the reaction must proceed by transferring energy in integer multiples of phonon energies.

The conditions for observing the resonance effect--for resolving the rate maxima--may be derived as follows. Let us imagine that we have, among others, two modes separated in frequency by an amount  $\omega_1 - \omega_2 = \Delta\omega$ . Then the maximum at

$$\epsilon_a = \sum_{\mu \neq 1,2} n_\mu \omega_\mu + n_1 \omega_1 + n_2 \omega_2 \quad (\text{III.A.4.2})$$

is separated from its nearest neighbor

$$\epsilon_b = \sum_{\mu \neq 1,2} n_\mu \omega_\mu + (n_1 + 1)\omega_1 + (n_2 - 1)\omega_2 \quad (\text{III.A.4.3})$$

by an amount  $\Delta\omega$ . Thus if  $\Delta\omega$  is a typical spacing between vibrational frequencies in the molecule, so that  $(\Delta\omega)^{-1}$  is a coarse-grained density of vibrational states, then the resonances in the reaction rate as a function of energy gap are separated on average by  $\Delta\omega$ . If the typical mode frequency is  $\bar{\omega}$  and a typical linewidth is  $\bar{\gamma}$ , then we may approximate the widths of the resonances from Eq. (III.A.4.1) as

$$\Gamma = \sum_{\mu} n_\mu \gamma_\mu \approx \bar{\gamma} \sum_{\mu} n_\mu, \quad (\text{III.A.4.4})$$

but since

$$\epsilon = \sum_{\mu} n_\mu \omega_\mu \approx \bar{\omega} \sum_{\mu} n_\mu, \quad (\text{III.A.4.5})$$

we obtain

$$\Gamma \approx \bar{\gamma} \frac{\epsilon}{\bar{\omega}}. \quad (\text{III.A.4.6})$$

Thus the condition for observing quantum resonances in the reaction rate is

$$\Delta\omega \leq \bar{\gamma} \frac{\epsilon}{\bar{\omega}}, \text{ or}$$

$$\frac{\Delta\omega}{\bar{\gamma}} \leq \frac{\epsilon}{\bar{\omega}}. \quad (\text{III.A.4.7})$$

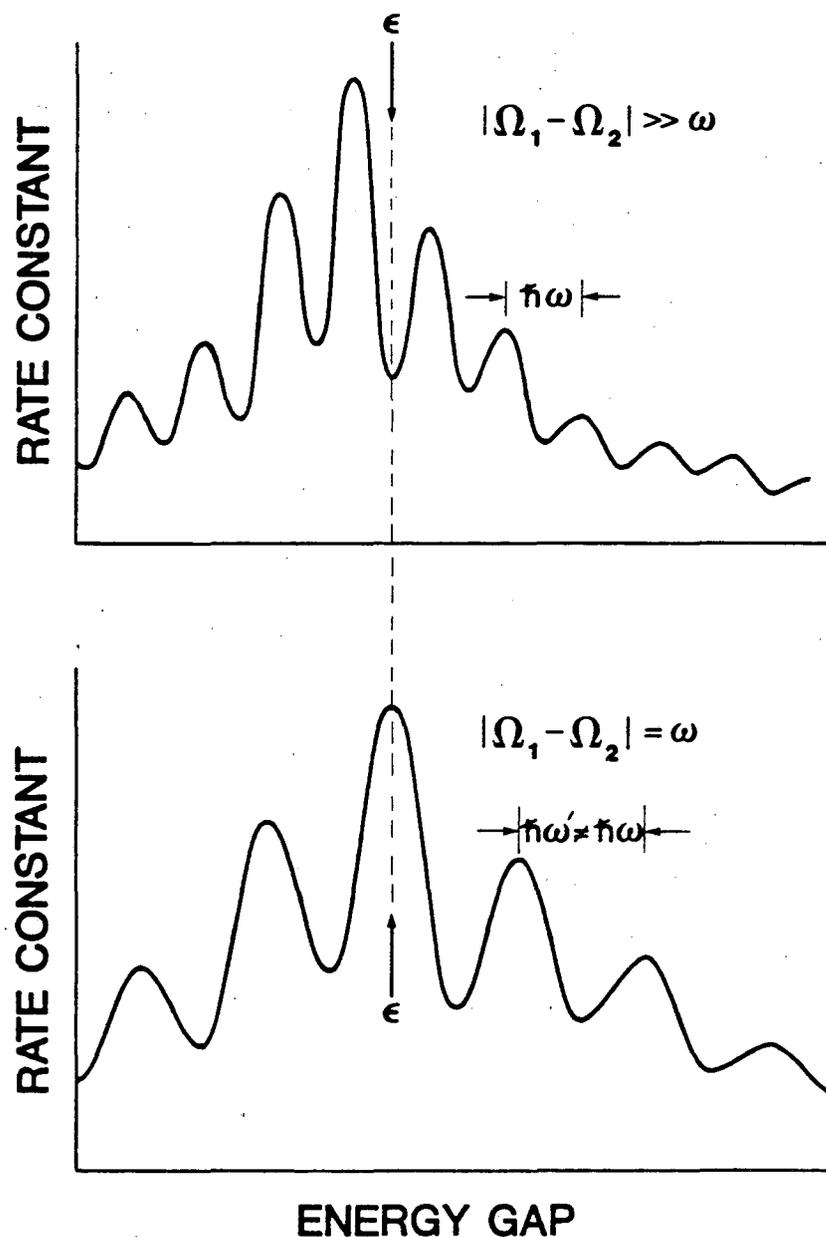
This condition for observing the resonance effect may be compared with experiment in the case of the DeVault-Chance reaction. Goldstein<sup>63</sup> has used the theory of Appendix D and Ref. 58 to show that the temperature dependent kinetics of this reaction are consistent with coupling to modes of average frequency  $\omega \approx 220 \text{ cm}^{-1}$ ; this conclusion is independent of whether one assumes that only one mode or a large number of modes are involved. The total coupling to these modes is strong,  $S \approx 6$ , and by the arguments of Section III.A.1 this coupling constant is consistent with the 0.01 *nm* motions observed crystallographically<sup>64</sup> upon oxidation or reduction of cytochrome *c*. This analysis of the data was confirmed by the measurement of a "charge-transfer band," which is a radiative transition between the same two electronic states ordinarily connected by a radiationless matrix element. As first predicted by Hopfield,<sup>65</sup> the parameters of this band can be used to independently test the model Hamiltonian used in the interpretation of the temperature-dependent kinetics. Goldstein<sup>63</sup> and Goldstein and Bearden<sup>66</sup> found that theory and experiment agree both in the position and total cross-section (which is very small;  $\sigma \approx 10^{-23} \text{ cm}^2$ ) of the charge-transfer absorption.

The critical region of the vibrational spectrum in the case of the DeVault-Chance reaction thus appears to be  $\bar{\omega} \approx 220 \text{ cm}^{-1}$ , while the energy gap for this reaction is known to be  $\epsilon \approx 0.46 \text{ eV}$ . If we look at published Raman spectra of cytochrome *c* and other heme proteins,<sup>67,68</sup> this region consists of about one mode every twenty to fifty wavenumbers, so that  $\Delta\omega \approx 30 \text{ cm}^{-1}$ . Thus the quantum resonances in the reaction rate would be resolved if the natural linewidth  $\bar{\gamma} \leq 2 \text{ cm}^{-1}$ , or equivalently if the

vibrational lifetimes were  $\tau_v = 1/2\pi\gamma \geq 2.5 \times 10^{-12} s$ , which is not unreasonable.<sup>69</sup> Qualitatively then, the parameters of the DeVault-Chance reaction are such that the quantum resonances should be just barely observable.

The observability of the quantum resonances is, however, a delicate phenomenon. If there are many more modes coupled to the reaction than are visible in the resonance Raman spectrum then the effect will be significantly reduced. On the other hand, just as biological molecules are often assumed to have structures which are ideally suited to their tasks, it is likely that whatever effects the vibrational spectrum may have are also optimized in real biological systems. But what function could the quantum resonances serve? The answer to this question brings us back to the exchange among Jordan,<sup>70</sup> Jehle,<sup>71,72</sup> and Pauling and Delbrück<sup>73</sup> which was mentioned in Chapter I.

When two molecules participate in a reaction--whether the two molecules are electron donor and acceptor or enzyme and substrate--the modes to which the reaction is coupled involve both molecules. Because the molecules are bound, however, the vibrational modes of the donor/acceptor or enzyme/substrate complex are not simply those of the two molecules alone, but rather some new modes which take account of the coupling that occurs upon binding. To make this point concrete, consider a small substrate molecule with two relatively high frequency modes at  $\Omega_1$  and  $\Omega_2$  bound to an enzyme molecule with one low frequency mode at  $\omega$ . If  $|\Omega_1 - \Omega_2| \approx \omega$ , then small anharmonic terms in the Hamiltonian will cause large shifts in the vibrational frequencies of the enzyme-substrate complex, so that the frequencies will be very different depending on whether or not this resonance condition (or some higher order resonance condition  $|n_1\Omega_1 - n_2\Omega_2| \approx m\omega$ ) is met. Under conditions where quantum resonances are observable in the reaction rate, this vibrational interaction between enzyme and substrate will lead to selective enhancement of the reaction rate for some substrate molecules over others, as illustrated in Fig. III-12.



**FIGURE III-12** Quantum resonances and kinetic specificity. Above, the rate constant as a function of energy gap for a substrate whose vibrational frequencies are not in resonance with those of the enzyme; the rate maxima are spaced simply by the enzyme frequency, and the actual energy gap corresponds to a minimum rate: the reaction is inhibited. Below, same as above but for a resonant substrate; now the spacing among maxima reflects the vibrational frequency shifts which result from anharmonic coupling among the modes, and the rate is enhanced.

If the coupling of electronic states to vibrational modes in biological molecules is not random--that is, if not all modes are equally strongly coupled or if there is some pattern to the frequencies of the coupled modes--then the quantum mechanical

resonance effect can, by the arguments of Fig. III-12, give rise to tremendous specificity in the rates of "correct" vs. "incorrect" reactions involving a given enzyme. This is a modern version of the effect first suggested by Jordan<sup>70</sup> and so clearly rejected by Pauling and Delbrück.<sup>73</sup> An important difference, however, is that the effect considered here is a *kinetic* one, while Jordan sought to understand an *equilibrium* effect on binding energies.

What we have seen is that the parameters of the DeVault-Chance reaction, as an example, are consistent with the occurrence of the kinetic effect, while the possibility of equilibrium effects has not been addressed. If such kinetic effects indeed occur they will have widespread implications for our understanding of specificity in biological processes. Sections III.B.1-4 are concerned with demonstrating the existence of the resonance effect in the interactions of heme proteins with small ligands, a system exhibiting a number of anomalous kinetic features which may be interpreted as arising from the quantum resonances.

The resonant dependence of the electronic self-energy on vibrational frequency is obtained whether the rate of the electronic transition is fast or slow compared to the vibrational relaxation rate. If the transition rate is fast compared to vibrational relaxation, however, a new quantum effect appears, namely electronic coherence. The method used here for describing the approach of a quantum system to equilibrium is that the "proper subsystem," in this case the electrons and phonons taken together, is coupled to a heat bath. If the coupling to the bath did not occur, the electron-phonon system would evolve in a completely deterministic fashion and would obey conservation of energy. Furthermore, to the extent that the harmonic approximation is valid for the phonons, there would be no sharing of energy among the various vibrational modes of the system, so that the molecules would not even reach an internal "temperature." In particular, in the absence of coupling to heat bath, if the system starts in a pure state it must stay in a pure state--there is no loss of quantum mechanical

coherence in the absence of vibrational relaxation.

The fact that coherence is conserved on a time scale short compared to vibrational relaxation does not mean that the system will not *appear* to be evolving toward an equilibrium state. Consider again a model with two electronic states and one phonon mode, as illustrated in Fig. III-8. We have already seen that this system collapses to an infinite set of two level systems, one for each vibrational state. If the wavefunctions are labeled  $|\pm; n\rangle$  for the + or - electronic state and the  $n^{\text{th}}$  vibrational state, the time evolution of a system starting in state  $|+; n\rangle$  is

$$|\psi(t)\rangle = \cos(VF_n t)|+; n\rangle + \sin(VF_n t)|-; n + \epsilon/\omega\rangle, \quad (\text{III.A.4.8})$$

where  $F_n$  is the matrix element  $|\langle n + \epsilon/\omega|F|n\rangle|$ , and the operator  $F$  is defined in Eq. (III.A.3.42). If we start in a superposition of vibrational states, the wave function becomes

$$|\psi(t)\rangle = \sum_n a_n(0) [\cos(VF_n t)|+; n\rangle + \sin(VF_n t)|-; n + \epsilon/\omega\rangle]. \quad (\text{III.A.4.9})$$

The time evolution of the electronic state population is therefore

$$P_+(t) = \text{Tr}_{ph} |\langle +|\psi(t)\rangle|^2 = \sum_n |a_n(0)|^2 \cos^2(VF_n t), \quad (\text{III.A.4.10})$$

where  $\text{Tr}_{ph}$  denotes a trace over phonon degrees of freedom. What we see is that, although the electronic degree of freedom is not in contact with the heat bath, it seems to exhibit a quasi-irreversible decay, since in general the many oscillation frequencies  $VF_n$  are incommensurate. This decay occurs at a "rate" roughly given by the width of the distribution of oscillation frequencies, which is proportional to  $V$  and dependent upon the initial distribution over vibrational states.

The dynamics of the electronic states for times comparable to or shorter than the vibrational relaxation time are therefore misleading. If we observe the predicted decay, we might be tempted to fit a rate constant to its time course and to interpret this rate constant and its temperature dependence in terms of simple perturbation theory. We would then expect the rate to be proportional to  $V^2$  according to the Golden rule; in fact the rate will be proportional to  $V$ . We might imagine that the apparent irreversibility is a manifestation of thermodynamic principles and the approach to equilibrium; in fact the decay is only a consequence of the ensemble in which the measurement was begun and would resolve itself into oscillations if we could monitor the individual vibronic states rather than just the average electronic population.

It is important to understand that these results are not at all mysterious. By definition we have agreed that the temperature of the system and all its thermodynamics are enforced by coupling of the vibrational modes to a heat bath formed, for example, by the solvent. At sufficiently short times, the matrix elements which give rise to this coupling cannot have a significant influence on the dynamics of the system. As a result, the molecule does not "know" the temperature of the world, and cannot possibly relax toward it. The rate of the reaction then does not depend on the temperature of the world, but on the initial ensemble in which the system was prepared, for example by a short optical pulse; of course this pulse draws the molecules out of a thermal ensemble but leaves them in a non-thermal ensemble. The nature of the initial flash thus should influence the time course of the electronic dynamics, and experiments with different flash conditions can be compared only very carefully. These problems in interpreting experiments in the picosecond regime are the topic of Section III.C.1.

As noted in Section III.A.1, the primary events of photosynthesis include at least one reaction which competes with vibrational relaxation. Consequently the coherence effects predicted here should be observable, although their detection may prove

difficult. Sections III.C.2-4 discuss the evidence that one consequence of quantum coherence has in fact been found in this biological system. Finally, there is evidence that in at least some biological polymers the vibrational relaxation times can be nanoseconds or more<sup>41</sup>; in these systems quantum coherence should be accessible even for "ordinary" reactions which are not triggered by photon absorption but rather by ligand binding or other chemical events. Sections III.D.1-3 consider the implications of such coherent behavior, and in particular the possibility that this effect can lead to macroscopic quantum effects as required to understand quantum-limited measurement in the sensory systems.

## **B. Resonance in the slow regime: Theory vs. experiment in myoglobin**

### *1. Vibrational lineshapes and non-exponential kinetics*

One of the central results of a rigorous quantum mechanical theory of molecular dynamics is the existence of resonances in the dependence of kinetic parameters on the vibrational frequencies of the molecules. Thus the rate of transition between two electronic states of a molecule is maximum whenever the energy difference between the two states corresponds to an integer number of phonons, that is  $\epsilon \approx N\hbar\omega$ . This effect becomes less significant when a large number of modes are coupled to the electronic transition, but persists for the case of a few modes even at high temperatures ( $k_B T \gg \hbar\omega$ ; cf. Appendix D). It is thus a true quantum effect which can exist even in the nominally semi-classical regime of large phonon occupation numbers.

The existence of these resonances implies that a distribution of frequencies can lead to a broad distribution of transition rates. Thus if the rate is  $k(\omega)$  and the probability of a molecule having a frequency  $\omega$  is  $P(\omega)$ , then the average time course of the transition from one electronic state to the other is given by<sup>74</sup>

$$F(t) = \int d\omega P(\omega) e^{-k(\omega)t}. \quad (\text{III.B.1.1})$$

We may approximate the resonant dependence of the reaction rate on frequency as a sum of Gaussians (cf. Appendix D); if one of these terms dominates,

$$k(\omega) = k_0 \exp\left[-\frac{(\omega - \omega_r)^2}{2\sigma^2}\right]. \quad (\text{III.B.1.2})$$

Similarly, the inhomogeneous lineshape  $P(\omega)$  is generally Gaussian in form (see below), so that

$$P(\omega) = \frac{1}{[2\pi(\Delta\omega)^2]^{1/2}} \exp\left[-\frac{(\omega - \omega_0)^2}{2(\Delta\omega)^2}\right], \quad (\text{III.B.1.3})$$

where  $\Delta\omega$  is the inhomogeneous linewidth; for simplicity we may assume that the resonant frequency  $\omega_r$  coincides with the center frequency  $\omega_0$ . The time course  $F(t)$  now becomes

$$F(t) = \int \frac{d\omega}{[2\pi(\Delta\omega)^2]^{1/2}} \exp\left[-\frac{(\omega - \omega_0)^2}{2(\Delta\omega)^2} - k_0 t \exp\left[-\frac{(\omega - \omega_0)^2}{2\sigma^2}\right]\right]. \quad (\text{III.B.1.4})$$

The quantum resonances thus have a unique experimental signature: non-exponential decay of the sort predicted by Eq. (III.B.1.4). These decays are presented in Fig. III-13 for various values of the critical parameter  $\Delta\omega/\sigma$ . In the long time ( $k_0 t \gg 1$ ) limit, the integral in Eq. (III.B.1.4) may be evaluated by a saddle-point approximation, giving

$$F(t) \approx e^{-\left(\frac{\sigma}{\Delta\omega}\right)^2 [1 + 2\ln(\Delta\omega/\sigma)]} \frac{(k_0 t)^{-(\sigma/\Delta\omega)^2}}{[2\ln\{k_0 t (\Delta\omega/\sigma)^2\}]^{1/2}}, \quad (\text{III.B.1.5})$$

which is a power-law decay modified by a logarithmic term.

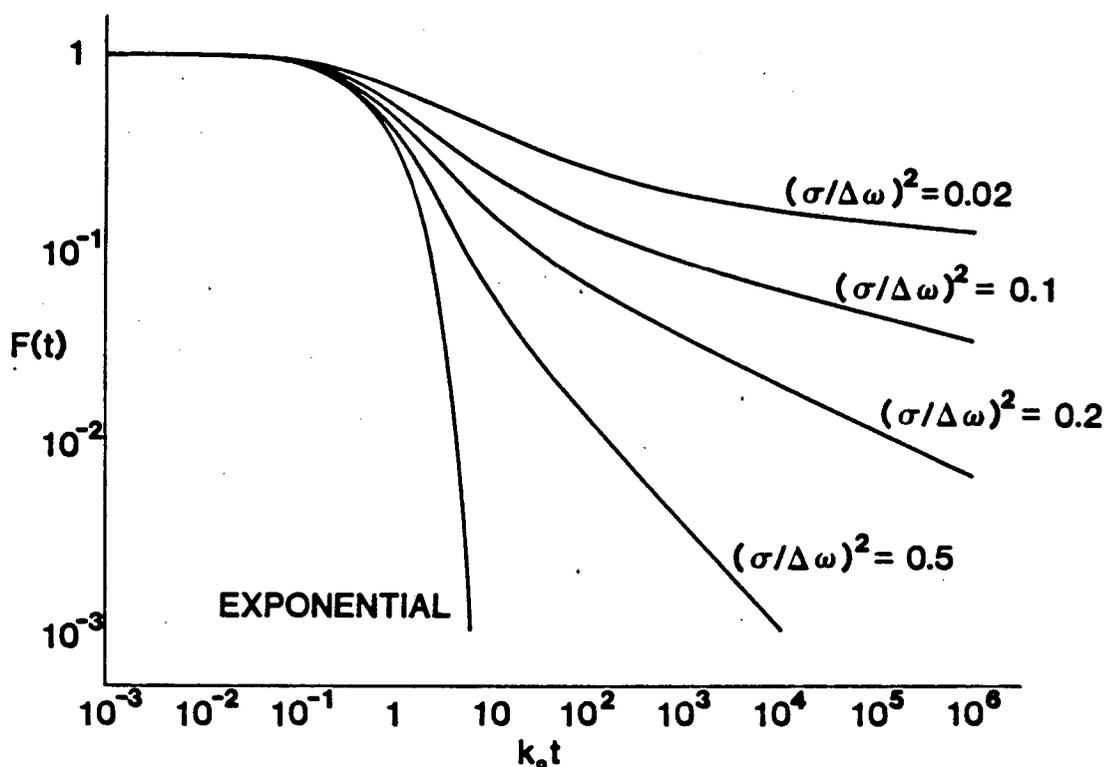


FIGURE III-13 Non-exponential decays generated by quantum resonances. The time course of electronic transitions in the single-resonance approximation of Eq. (III.B.1.4).

Power law decays have been observed in the ligand binding reactions of the heme proteins in frozen solution,<sup>27</sup> as described in Section III.A.1. The arguments of this and the following sections test the hypothesis that the observed non-exponential behavior in this biological systems in fact reflects the quantum mechanical resonance effect.

The detailed spectroscopic and structural studies on myoglobin which have been done in both the ligand bound ( $(Mb \cdot CO)_{S=0}$ ) and photodissociated ( $(MbCO)_{S=2}^*$ ) states allow us to define, at least approximately, the critical vibrational modes which are coupled to the reaction and to estimate their coupling constants:

- [1] The Fe-C stretching mode<sup>75</sup> ( $\omega \approx 500 \text{ cm}^{-1}$ ) changes frequency, since the Fe-C bond is broken upon photodissociation, and the equilibrium bond length changes by<sup>76</sup>  $\delta x \leq 0.005 \text{ nm}$  upon ligand binding.

- [2] The C-O stretching mode ( $\omega \approx 2000 \text{ cm}^{-1}$ ) changes frequency<sup>77</sup> by  $\approx 200 \text{ cm}^{-1}$ , and the equilibrium bond length may also change.
- [3] The Fe-His stretching mode<sup>78</sup> ( $\omega \approx 220 \text{ cm}^{-1}$ ) changes length by<sup>76</sup>  
 $\delta x = 0.004 \pm 0.002 \text{ nm}$ .

All of these modes, however, have  $\hbar\omega \gg k_B T$  throughout the temperature range of interest, and thus cannot contribute to the temperature dependence of the reaction time-course. Further, as discussed in Appendix D, the effect of these modes can be summarized as a renormalization of the matrix element and energy gap, and with this renormalization the high frequency modes need not be treated explicitly.

- [4] The "F" alpha helix in Mb appears to move in a collective mode, being rotated and possibly stretched when a ligand binds.<sup>79</sup> An average atom in the F helix is displaced by  $\approx 0.03 \text{ nm}$ .

From the analysis presented at the end of Sect. III.A.2, we expect that these collective modes have frequencies  $\approx 20 \text{ cm}^{-1}$ , since the F helix is  $\approx 1 \text{ nm}$  long. Further, the rather small displacement of  $\approx 0.03 \text{ nm}$  corresponds to a coupling constant of  $S \approx 50$ , again using the results of Section III.A.2. Brillouin scattering experiments on alpha-helical polymers<sup>41</sup> demonstrate that these collective modes have very small damping constants,  $\gamma \approx 10^9 \text{ s}^{-1}$ , or natural linewidths  $\gamma/2\pi \approx 0.005 \text{ cm}^{-1}$ . These narrow linewidths suggest that the resonance effect in reactions coupled to the collective modes of an alpha helix will be very large, which is precisely what we require to account for the extreme non-exponentiality of the ligand binding reactions in myoglobin. As a first approximation, I will therefore consider a very simple model, as in Fig. III-14, in which the ligand binding reaction is strongly coupled only to a single low frequency, narrow linewidth vibrational mode.

Calculation of the reaction time course in a single mode model requires estimates for six parameters: The mode frequency, natural linewidth, and coupling constant; the energy gap  $\epsilon$  which separates the electronic states and the matrix element  $V$  which

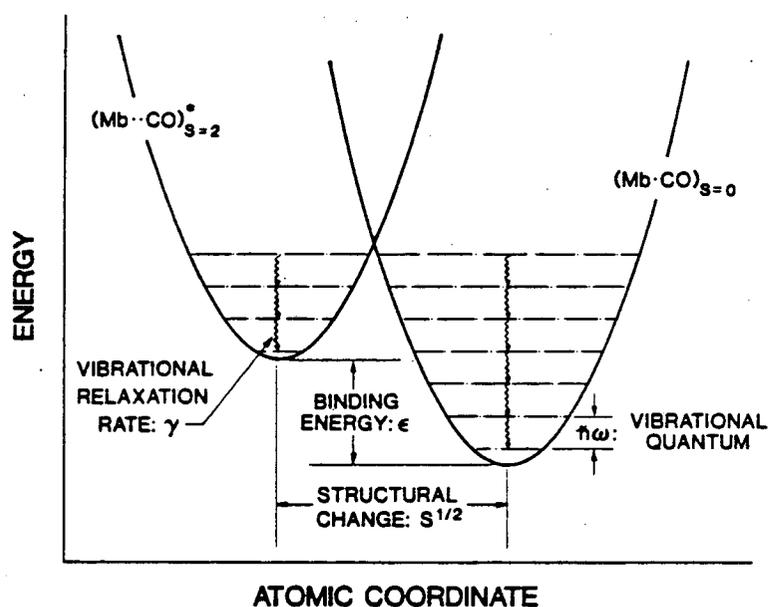


FIGURE III-14 A simple single-mode model Hamiltonian for carbon monoxide binding in myoglobin; parameters such as the vibrational quantum and the energy gap are not shown to scale.

connects them; and the inhomogeneous vibrational linewidth  $\Delta\omega$ . Figure III-15 illustrates the rate constant as a function of frequency and temperature with the first three parameters chosen in accord with the estimates given above; the remaining parameters are chosen to match the experiments on the temperature-dependent time course of the ligand binding reaction, also shown in Fig. III-15 in comparison with the theory. The fit determines  $\epsilon \approx 0.3 \text{ eV}$ , which is in good agreement with the interpretation of the kinetic data,<sup>27</sup>  $V \approx 10^{-4} \text{ eV}$ , which is in accord with the estimates of Hopfield and co-workers based on the magnetic properties of the heme iron,<sup>80</sup> and as noted below the model then requires only a small amount of inhomogeneous broadening to generate the extreme non-exponential behavior which is in fact observed. It is clear from Fig. III-15 that reasonable agreement is obtained even with this simplest model.

One important consequence of the quantum resonances is the existence of anomalous isotope effects. The ligand binding reaction in heme proteins involves changes in the frequency of a number of high-frequency modes, as noted

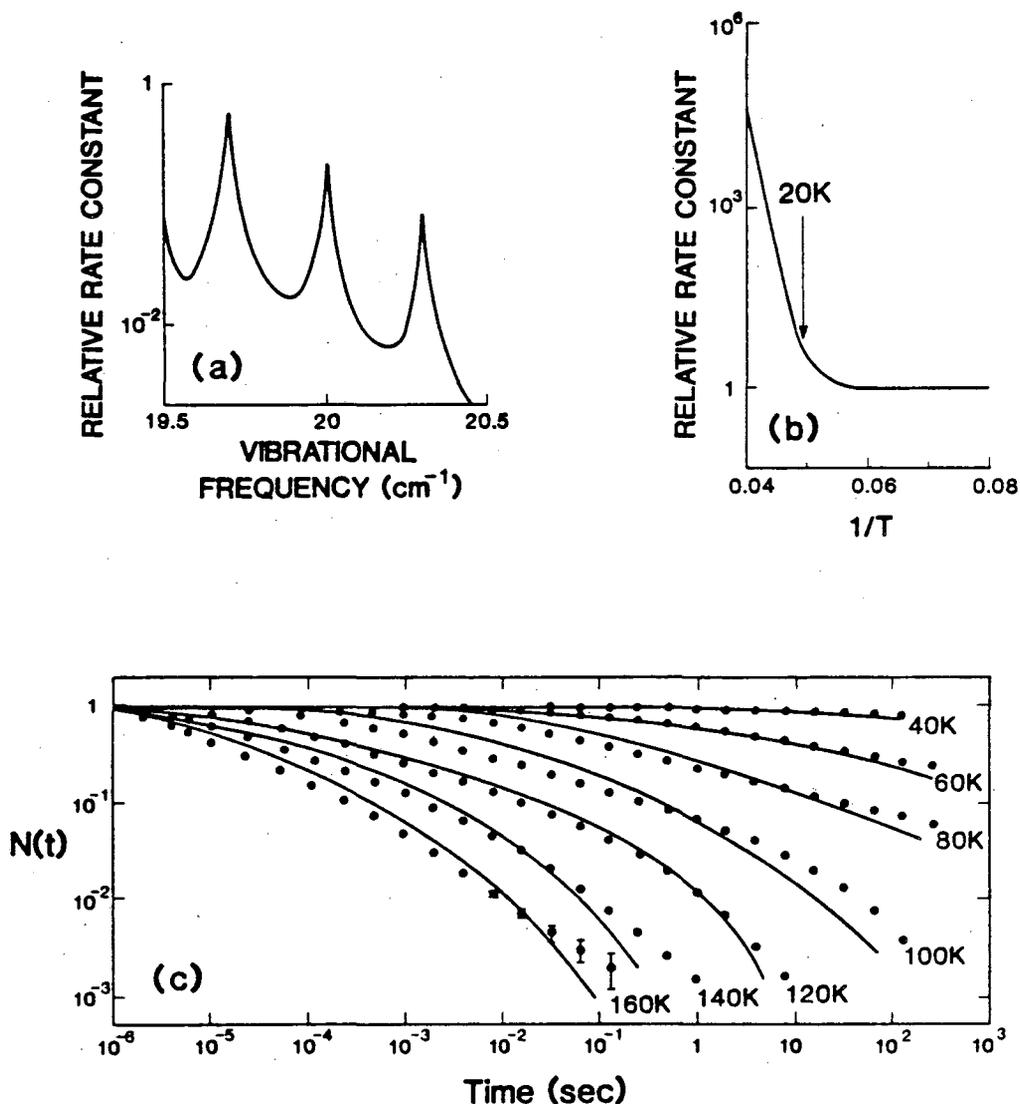


FIGURE III-15 Ligand binding in heme proteins--Theory. (a) Rate constant as a function of vibrational frequency; the variations in rate are more than three orders of magnitude for a frequency range of only one wavenumber. (b) Rate constant, for fixed energy gap and vibrational frequency, as a function of temperature; transition from temperature-independence to Arrhenius behavior at  $\approx 30$  K as observed experimentally (cf. Fig. III-5). (c) Reaction time course, calculated from Eq. (III.B.1.1), in comparison with experiment; data as shown in Fig. III-5. Details of the calculations are given in Appendix D.

above.<sup>75,77,78</sup> Together, these frequency changes amount to a difference in zero-point energy of  $\approx 200 \text{ cm}^{-1}$  between the two electronic states. When the carbon or oxygen atom is isotopically substituted, the effective mass associated with these modes changes by  $\approx 10\%$ , so that the frequencies and energies change by 5% or  $\approx 10 \text{ cm}^{-1}$ .

This is close to the separation between resonances, so that isotopic substitution will move the system from one resonance to the next, resulting in a large change in the reaction rate. Further, this change will not depend in any simple way on the magnitude of the mass changes, and may even be non-monotonic if additional mass changes move the energy gap through the successive resonances. In fact, experiments in myoglobin demonstrate that at very low temperatures such non-monotonic isotope effects do occur.<sup>81</sup> As the temperature is increased, this effect becomes smaller, which is consistent with the fact that the quantum resonances broaden at higher temperature.

These anomalous isotope effects have previously been interpreted in terms of quantum mechanical tunneling of the Fe-CO system; the non-monotonicity was taken as evidence that internal motions of the CO, in addition to rigid motion toward the iron, occurs during the rebinding.<sup>81</sup> It had been suggested earlier<sup>28,30</sup> that the temperature independent rebinding rate at low temperature reflects such tunneling processes, with the temperature-dependent rebinding at  $T > 30 K$  reflecting classical "over the barrier" motion. In this picture, the decrease in isotope effect with increasing temperature measures the decrease in the contribution of tunneling relative to the classical motion.<sup>81</sup> Is this interpretation correct, or does the anomalous isotope effect provide evidence of quantum resonances, as suggested here?

The cross-over from temperature independent reaction rates to activated behavior occurs as the thermal energy results in population of the excited vibrational states of the system. This is true whether we use a semi-classical approximation, an intuitive quantum analysis in terms of stimulated and spontaneous phonon emission, or a rigorous multi-phonon theory, as was emphasized in Section III.A.1. The fact that this cross-over occurs at  $\approx 30 K$  in myoglobin means that some molecular vibration of frequency  $\omega \approx 20 \text{ cm}^{-1}$  must be coupled to the reaction, and this is what we find for the fit shown in Fig. III-15.

In contrast, the local motions of the Fe-CO system have frequencies ranging from  $200\text{ cm}^{-1}$  to  $\approx 2000\text{ cm}^{-1}$ , so that all these modes are trapped in their ground states throughout the temperature range of interest. Indeed it may be shown that if, for example, the rocking motion of the CO molecule in the heme pocket had a frequency of  $\approx 20\text{ cm}^{-1}$  then the zero-point motion associated with this mode alone would result in random rotations of the molecule by more than ninety degrees. Similarly such a low frequency for the Fe-C stretching mode would result in zero-point motions on the order of several angstroms; both these numbers are in clear disagreement with the observability of the CO molecule in a crystallographic analysis<sup>82</sup> and with the existence of well-defined structure in the X-ray absorption spectrum of the iron atom.<sup>76</sup> Thus if we wish to associate the isotope effect with tunneling of the Fe-CO system then we cannot possibly account for its temperature dependence.

This argument may be carried further. Analysis of the X-ray absorption spectra before and after flash photolysis<sup>76</sup> demonstrates that the relative motion of the iron and carbon atoms is  $0.005\text{ nm}$  or less. Knowing the frequency of the Fe-C stretching mode<sup>75</sup> and taking its mass as the reduced mass of the Fe atom and CO molecule, one can calculate the coupling constant  $S$  which corresponds to the observed displacement; the result is  $S \leq 1$ . At low temperatures, the requirement that the CO molecule tunnel through this distance suppresses the reaction rate by the Franck-Condon factor  $e^{-S}$ , and it is the relation  $S \approx m^{1/2}$  which accounts for the isotope effect on tunneling. But if  $S \leq 1$ , it is impossible to generate isotope effects of order twenty percent, as observed.<sup>81</sup> In summary, the observed temperature-dependent anomalous isotope effects in the myoglobin-CO reaction cannot be accounted for in terms of molecular tunneling. At the same time, they have a natural explanation in terms of the same quantum resonances used here to account for the non-exponential kinetics, and thus provide excellent confirmation of this hypothesis.

The central feature of the description in terms of quantum resonances is the requirement for a very low frequency ( $\approx 20 \text{ cm}^{-1}$ ) narrow linewidth ( $\gamma/2\pi \approx 0.005 \text{ cm}^{-1}$ ) mode with very strong ( $S \approx 50$ ) coupling to the reaction. The extreme non-exponential behavior observed experimentally can be then generated by very small amounts ( $\Delta\omega \approx 1 - 10 \text{ cm}^{-1}$ ) of inhomogeneous broadening. Several questions arise from these results:

- [1] Is there independent evidence for the narrow linewidth of the protein breathing mode?
- [2] Can we predict the conditions required for non-exponential decays to be observed? In particular, why is this behavior restricted to frozen solutions,<sup>27</sup> rather than simply low temperatures?
- [3] What is the mechanism for inhomogeneous broadening of the low-frequency mode?

To approach these questions, or more generally to provide quantitative tests of the theory presented here, the parameters of the model Hamiltonian must be determined by independent experiments.

## *2. Direct measures of vibrational lineshape*

In a perfect, regular crystal, all vibrational modes are extended, in accord with Bloch's theorem. In a strongly disordered system, such as an amorphous solid or a large molecule, we expect the vibrational modes to be localized,<sup>83</sup> or non-propagating. This is not to say, however, that the modes are so strongly localized that they involve only motions of a few atoms which are in van der Waals contact. Rather there is a whole spectrum of modes in a macromolecule, ranging from relatively high frequency modes associated with individual bonds to relatively low frequency modes associated with larger scale "breathing" motions in which large portions of the molecule participate.

The high frequency local modes and the low frequency breathing modes are very different in character. Individual interatomic bonds are highly anharmonic, being described approximately by the Morse potential.<sup>84</sup> For these modes, the first vibrational excited state--which is reached by infrared absorption--is well above the energy range in which the potential surface can be considered parabolic. For the low frequency modes, on the other hand, the stretching of any individual bond is very small, since the strain is distributed over the large number of bonds involved in the "breathing". Thus, for the first few excited vibrational states (and possibly more) the low frequency modes can be treated as harmonic.

The low and high frequency modes also differ in the extent of their coupling to the heat bath. We expect that the heat bath in the biological systems of interest here is formed mostly by the solvent. Highly localized motions of atoms in the interior of a macromolecule will therefore not couple very strongly to the bath, and can relax only by anharmonic coupling to other modes. The low frequency modes, on the other hand, will contribute to the motions of most of the atoms in the molecule, including those at the surface, and through these surface groups the low frequency modes can interact with, and be damped by, the solvent.

When we dissolve a molecule in solution, the vibrational frequencies of the molecule shift relative to their gas-phase values. This solvent shift is exerted by the molecules in the "solvent shell," which for a protein like myoglobin consists of about fifty to one hundred water molecules which are ordered on the protein surface and hence visible crystallographically.<sup>82,85-87</sup> Molecules will exchange between the solvent shell and the bulk solvent phase, however, so that the occupancy of the solvent shell and the magnitude of the solvent shift will be stochastic quantities. It is this stochastic occupancy of the solvent shell which gives rise to much of the inhomogeneous broadening of vibrational lines in solvated molecules.<sup>69</sup>

The time scale for fluctuations in solvent shell occupancy are, at room temperature,<sup>88</sup> on the order of nanoseconds. This is long compared to vibrational relaxation, but short compared to most chemical reactions. Thus when one looks at the vibrational spectrum with infrared or Raman spectroscopy--where the measurement time is set by the relaxation rate--we see an essentially static distribution of vibrational frequencies; because a large number of solvent molecules contribute to this effect the distribution is approximately Gaussian. When one examines the time course of chemical reactions, however, the measurement time is so much longer than the time scale of the frequency fluctuations that we see only an average reaction rate, and an exponential time course for the transition.

When the solvent is frozen, the time required for exchange between the solvent shell and bulk phase becomes very long, and the distribution of frequencies is quite literally frozen on the chemical time scale. If the rate constant for a reaction is strongly dependent on vibrational frequency, as in the case of quantum resonances, this static distribution of frequencies is reflected as a static distribution of rate constants, and the theory of non-exponential decays proceeds as presented above.

We thus arrive at the following physical picture of the origin of non-exponential behavior in the heme proteins: The mode which is most strongly coupled to the transitions among electronic states is of very low frequency and therefore involves motions of a large fraction of the protein, including groups at the molecular surface. Through the surface groups the mode interacts with the solvent, giving rise both to vibrational relaxation and a stochastic solvent shift of the vibrational frequency. In frozen solvents this stochastic shift is stationary on the time scale of the transition, so that we can observe the effects described in the previous section, namely a distribution of reaction rates resulting from the quantum resonances.

The distribution of solvent shell occupancy, and hence of reaction rate, is not an equilibrium distribution in the frozen state. As in the case of glasses, equilibration of

the solvent structure is blocked by solidification, so that the particular non-exponential time course of the transition must reflect the conditions of freezing and not simply the conditions of the measurement. Experiments on myoglobin in which both temperature and pressure were manipulated<sup>89</sup> confirm this prediction.

One critical test of this explanation of the non-exponential behavior is to directly observe the inhomogeneous broadening of the low-frequency mode, and if possible to demonstrate that the number of relevant modes is sufficiently small for the quantum resonances to be expressed. The spectrum of low frequency modes may be probed through their effect on the vibrational relaxation, and hence on the infrared spectra of a high frequency mode. In particular we can compare the theory to experiments<sup>77,90,91</sup> on the infrared spectrum of the stretching mode ( $\approx 2000 \text{ cm}^{-1}$ ) of the carbon monoxide bound at the myoglobin active site.

Consider a high frequency mode with normal coordinate  $Q$  and frequency  $\Omega$ , and a low frequency mode with coordinate  $q$  and frequency  $\omega$ . In the absence of anharmonicity, the Hamiltonian has the form

$$\mathbf{H}_{\text{harmonic}} = \frac{1}{2}(\dot{Q}^2 + \Omega^2 Q^2) + \frac{1}{2}(\dot{q}^2 + \omega^2 q^2) + \Gamma \dot{q} + \mathbf{H}_{\text{hb}}, \quad (\text{III.B.2.1})$$

where  $\mathbf{H}_{\text{hb}}$  is the Hamiltonian of the heat bath and  $\Gamma$  is some generalized coordinate of the bath (cf. Section III.A.2); in accord with the discussion above, only the low frequency mode is directly coupled to the heat bath. If we are interested in the infrared absorption band of the high frequency mode, we must add to the Hamiltonian a term

$$\mathbf{H}_{\text{IR}} = \mathbf{H}_{\text{rad}} + Q \frac{\partial \bar{D}}{\partial Q} \cdot \bar{E}, \quad (\text{III.B.2.2})$$

where  $\mathbf{H}_{\text{rad}}$  is the Hamiltonian of the radiation field alone,  $\bar{E}$  is the electric field, and  $\bar{D}$  is the dipole moment of the molecule.

The lowest order anharmonicity which can enter is a cubic function of the two normal coordinates and there are four such terms:

- [1] A pure cubic anharmonicity of the high frequency mode itself,

$$H_{anharmonic}^{(1)} = \frac{a}{3!} Q^3. \quad (\text{III.B.2.3})$$

- [2] A pure cubic anharmonicity of the low frequency mode,

$$H_{anharmonic}^{(2)} = \frac{b}{3!} q^3. \quad (\text{III.B.2.4})$$

- [3] Anharmonic coupling between modes, linear in the high-frequency mode,

$$H_{anharmonic}^{(3)} = \frac{c}{3!} q^2 Q. \quad (\text{III.B.2.5})$$

- [4] Anharmonic coupling between modes, linear in the low frequency mode,

$$H_{anharmonic}^{(4)} = \frac{d}{3!} q Q^2. \quad (\text{III.B.2.6})$$

We shall see that each of these terms has a distinct effect on the infrared spectrum of the molecule. In order to calculate these effects, I again transform to a second quantized formulation.

The infrared absorption process involves only two states of the high frequency vibration, the ground and first excited states, and we may use fermion annihilation and creation operators  $c_0, c_0^\dagger, c_1, c_1^\dagger$  for the two states, respectively. In terms of these operators the full Hamiltonian may be written as

$$\begin{aligned} H = & \langle 0 | H | 0 \rangle c_0^\dagger c_0 + \langle 1 | H | 1 \rangle c_1^\dagger c_1 \\ & + \langle 1 | H | 0 \rangle c_1^\dagger c_0 + \langle 0 | H | 1 \rangle c_0^\dagger c_1 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2} \left( \Omega + \frac{d}{3! \Omega} q \right) (c_1^\dagger c_1 - c_0^\dagger c_0) + \left[ \frac{a}{2(2\Omega)^{3/2}} + \frac{cq^2}{3!(2\Omega)^{1/2}} \right] (c_1^\dagger c_0 + c_0^\dagger c_1) \\
 &\quad + \frac{1}{2} (\dot{q}^2 + \omega^2 q^2) + \Gamma \dot{q} + \mathbf{H}_{hb} + \frac{d}{3! \Omega} q + \frac{b}{3!} q^3 \\
 &\quad + \mathbf{H}_{rad} + \frac{1}{(2\Omega)^{1/2}} \frac{\partial \bar{D}}{\partial Q} \cdot \bar{E} (c_0^\dagger c_1 + c_1^\dagger c_0). \quad (\text{III.B.2.7})
 \end{aligned}$$

As discussed in Section III.A.3, we may introduce creation and annihilation operators for the low frequency mode which depend upon the state of the high-frequency mode:

$$\bar{a} = (q - q_{eq}) (\bar{\omega}/2)^{1/2} + i\dot{q} (\bar{\omega}/2)^{-1/2} \quad (\text{III.B.2.8a})$$

$$\bar{a}^\dagger = (q - q_{eq}) (\bar{\omega}/2)^{1/2} + i\dot{q} (\bar{\omega}/2)^{-1/2}, \quad (\text{III.B.2.8b})$$

where

$$q_{eq} = \frac{1}{b} \left\{ -\omega^2 + [\omega^4 - bd/2\Omega]^{1/2} c_1^\dagger c_1 + [\omega^4 - bd/6\Omega]^{1/2} c_0^\dagger c_0 \right\}, \quad (\text{III.B.2.9})$$

$$\bar{\omega}^2 = \omega^2 + bq_{eq}. \quad (\text{III.B.2.10})$$

With these definitions, the Hamiltonian becomes, up to an irrelevant constant,

$$\begin{aligned}
 \mathbf{H} &= \frac{\Omega}{2} (c_1^\dagger c_1 - c_0^\dagger c_0) + \bar{\omega} (\bar{a}^\dagger \bar{a} + 1/2) + i\Gamma (\bar{\omega}/2)^{1/2} (\bar{a}^\dagger - \bar{a}) + \mathbf{H}_{hb} \\
 &\quad + \frac{b}{3!} (\bar{a}^\dagger + \bar{a})^3 + \left[ \frac{a}{2(2\Omega)^{3/2}} \right. \\
 &\quad \left. + \frac{c}{3\bar{\omega}(2\Omega)^{1/2}} (\bar{a}^\dagger + \bar{a} + (\bar{\omega}/2)^{1/2} q_{eq})^2 \right] (c_1^\dagger c_0 + c_0^\dagger c_1) \\
 &\quad + (2\Omega)^{-1/2} \frac{\partial \bar{D}}{\partial Q} \cdot \bar{E} (c_0^\dagger c_1 + c_1^\dagger c_0) + \mathbf{H}_{rad}. \quad (\text{III.B.2.11})
 \end{aligned}$$

Equation (III.B.2.11) has the same form as discussed in Section III.A.3: it represents a two-level system coupled to a damped oscillator. The radiative absorption cross-section for this system can therefore be calculated using the methods of Appendix E.

All of the essential physics in the problem can be understood in a model where  $b = c = 0$ . In the case the Hamiltonian reduces to the simple model whose radiationless dynamics was discussed in Eq. (III.A.3.40) *et seq.*, and whose radiative interactions are discussed in detail in Appendix E. The two surviving anharmonic terms,  $a$  and  $d$ , have very different effects on the infrared spectrum. The term proportional to  $a$  results in a matrix element between the ground and first excited states of the high frequency mode, and therefore allows for vibrational relaxation. The term proportional to  $d$  may be added to the harmonic terms in the Hamiltonian for the high-frequency mode to give

$$H_{eff} = \frac{1}{2}[\dot{Q}^2 + (\Omega^2 + \frac{d}{3}q)Q^2], \quad (\text{III.B.2.12})$$

so that the low frequency mode gives rise to a frequency modulation of the high frequency mode. In semi-classical terms, the spectrum of the high frequency mode will therefore be split into side bands spaced by the frequency of the lower mode. This result persists in a lowest order treatment of the full quantum problem, but the simple pattern of equal spacing is broken in higher order, as discussed in Appendix E. Finally, each of the sidebands will be broadened by inhomogeneous broadening of the low frequency mode. These qualitative features of the spectrum, which will be insensitive to the details of the model Hamiltonian (*e.g.* the inclusion of higher-order anharmonic terms or a small number of other high frequency modes such as the Fe-C stretch) may be directly compared with the experiments cited above.<sup>77,90,91</sup>

First, the spacing between absorption bands is still a rough indicator of the frequency of the lower mode, which is apparently  $\approx 20 \text{ cm}^{-1}$ . This agrees with the interpretation of the kinetic data, which requires a mode of this frequency to be

strongly coupled to the motions (binding and unbinding) of the carbon monoxide as it sits in the interior of the protein.

Second, if a large number of protein vibrational modes were coupled to the motion of the carbon monoxide, we would not predict well resolved lines in the infrared spectrum but rather broad smears. This "washing out" of the sidebands is the same effect as the washing out of the resonances in the dependence of reaction rate on frequency. Thus the interpretation of non-exponential decays in terms of quantum resonances makes the unambiguous prediction that the results of anharmonic coupling will be multiple resolved lines rather than broad smears, and this prediction is confirmed by experiment.

Finally, the fact that the sidebands of the experimental spectrum are approximately Gaussian implies that the low frequency mode is inhomogeneously broadened. The existence of this broadening is essential to the interpretation of the non-exponential kinetics presented here, and the magnitude of the broadening observed in the infrared spectrum is in agreement with that required in the fit to the kinetic data.

The extant infrared spectra can thus be described by the same model Hamiltonian which describes the temperature dependence of the reaction rate and its distribution. Since this model Hamiltonian displays non-trivial quantum mechanical behavior, the infrared spectra confirm that the conditions for such behavior are met in a biological system. It would be attractive, however, if we could give an independent experimental test of the theory used in interpreting the infrared spectrum. In particular, it has been suggested that the resolved bands of the infrared spectrum may arise from multiple allowed conformations of the myoglobin-CO complex,<sup>77,90</sup> and this picture must be distinguished from that presented here.

If the multiple conformation model is correct, then each conformation corresponds to a slightly different environment of the CO and hence a slightly different force constant  $k$  for the stretching vibration; the effective mass of the mode,

however, will always be the reduced mass  $\mu$  of the C and O atoms, independent of conformation. Thus this model predicts that the frequencies of all the bands must change in proportion to  $\mu^{1/2}$ . In the anharmonic coupling model, the center frequency of the whole set of bands will be shifted by isotopic substitution, but the spacings among the bands (to first order) will not be affected. Thus the anharmonic coupling model predicts that isotopic substitution of the C or O atom will change the ratios of the band frequencies; the experiments<sup>77</sup> display changes of the sort predicted by the anharmonic coupling model, but these are not outside experimental error.

In addition to high resolution spectra, the two models can be distinguished by the behavior of the infrared spectrum during the ligand binding reaction. In the theory presented here, each individual molecule possesses an infrared spectrum with all sidebands present, and exhibits an exponential time course for the binding reaction. Different molecules have slightly different infrared spectra because of the inhomogeneous broadening, but drastically different reactions rates because of the resonance effect. Thus, as the reaction proceeds the shapes of the sidebands, as well as their integrated intensities, will change. The crucial point regarding the time dependence of the infrared spectrum is that it can be completely calculated within the framework of the present theory, and there is no need to introduce parameters beyond those used in the description of the static spectra. Experiments<sup>77</sup> show that the time course of the infrared spectrum is not simple, and displays a frequency dependence as predicted here, although this effect was interpreted by the experimenters in terms of the multiple conformation model. The multiple conformation model can make no quantitative prediction of the effect, and instead uses its existence to determine the rates of transition among the various conformations; these rates themselves must apparently be distributed so as to simulate power law decay.<sup>77</sup> Thus the conformational model, while giving an explanation of the extant experiments, is not tested by any of them. In addition, the conformational model does not predict any relations among the detailed

time course of the spectral changes and the origins of the non-exponential decay. In contrast the model Hamiltonians used here have already passed one test, in that the observable distributions of vibrational frequency and reaction rate have been quantitatively related, and further spectroscopic experiments, including the careful measurement of the time-dependent infrared spectrum, promise more stringent tests of the theory.

### *3. Vibrational spectra and intramolecular motions*

The vibrational spectrum of a protein can be probed not only by infrared spectroscopy, but also by techniques which are sensitive to atomic motions within the protein. Thus X-ray crystallographic Debye-Waller factors<sup>92</sup> provide estimates of the mean-square displacements of each (non-hydrogen) atom, while <sup>57</sup>Fe Mössbauer spectroscopy provides a related measure for the iron atom alone.<sup>93</sup> The large body of these data on myoglobin<sup>94-99</sup> have been interpreted almost exclusively in terms of a model illustrated in Fig. III-16a. In this picture the adiabatic potential surfaces of a single electronic state possess multiple minima, each corresponding to a "conformational substate" the dominant component of the protein dynamics is then hopping among these substates, rather than harmonic or near-harmonic vibrations as proposed here and schematized in Fig. III-16b.

It has been suggested that the substates used in the interpretation of the crystallographic data are related to the multiple allowed conformations postulated in relation to the infrared spectra,<sup>77</sup> and further that each substate has a different reaction rate, so that distribution over substates is responsible for the non-exponential kinetics.<sup>31</sup> The notion of conformational substates has now been used in the analysis of many other experiments<sup>100</sup> and appears to unify a large body of data.

If the multiple substates picture is correct, then the whole program of describing biological molecules in terms of simple model Hamiltonians will fail. This section is therefore devoted to a critical analysis of the extant crystallographic and Mössbauer

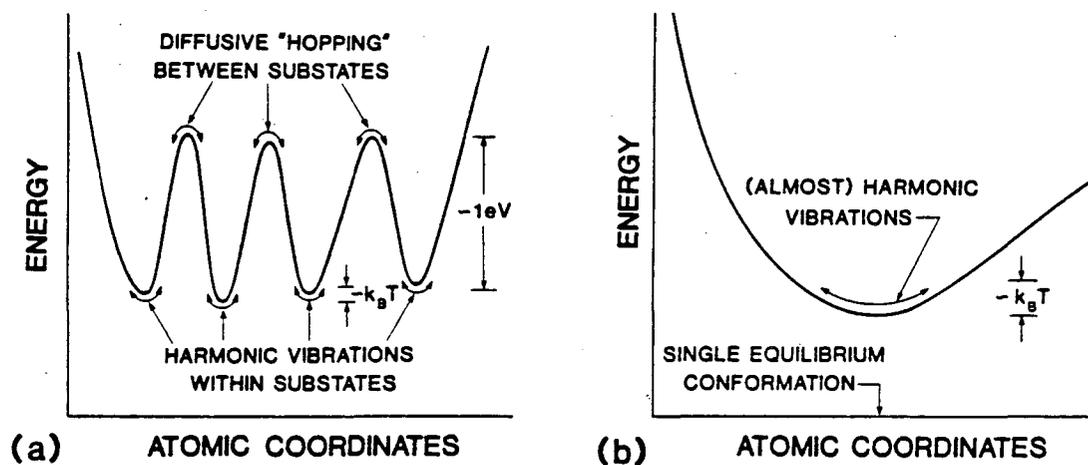


FIGURE III-16 Two models of protein dynamics. Left (a), the conformational sub-states model. Right (b), the quasi-harmonic model presented here.

experiments. I shall show that all of these data are in fact interpretable in terms of near-harmonic vibrations. Furthermore, the near-harmonic Hamiltonians allow the parameters extracted from the Mössbauer data to be compared with other probes of molecular dynamics, such as Raman spectroscopy. This analysis results in a “calibration” of the dimensionless coupling constants derived from the kinetic data against the actual atomic displacements in the protein.

I begin with the crystallographic data, which are unfortunately inconclusive. The standard Debye-Waller analysis of proteins<sup>86</sup> assumes that the position of each atom in the protein fluctuates according to an independent Gaussian distribution. In general we expect the distribution of positions to be given (in the classical limit) by Boltzmann’s principle as

$$P(\vec{x}) = Z^{-1} e^{-V(\vec{x})/k_B T}, \tag{III.B.3.1}$$

where the partition function  $Z$  is a normalization constant. For the rippled potential functions  $V(\vec{x})$  which are predicted by the substates model, this distribution cannot be Gaussian. In the harmonic model, on the other hand, the independently fluctuating

quantities are not the atomic coordinates but rather the normal coordinates; as a result the atomic displacements are not independent and there are concerted motions which correspond to the low frequency vibrational modes. Thus on either model the Debye-Waller analysis is not rigorously correct, and the results of such an analysis can therefore not be used to argue for either model. This conclusion also extends to those studies which have been done over a wide range of temperatures,<sup>94,95</sup> but which are nonetheless based on the same data analysis techniques.

The Debye-Waller factors do, however, provide a general scale for the mean-square displacements, and in the substates picture they determine the average displacement between adjacent substates<sup>94</sup>; this distance is  $d \approx 0.001 \text{ nm}$ . The other critical feature of the substates model is the magnitude of the barrier between states, which determines the temperature at which the states are "frozen in" this barrier has been estimated to be  $\Delta \approx 1 \text{ eV}$ .<sup>27</sup> In order for the substates interpretation of the kinetic data to be consistent, it must not be possible for the molecule to tunnel from one substate to the next, or else the distribution over substates would not be static on the time scale of the rebinding reaction and exponential decays would be observed (compare the discussion of solvent shell fluctuations in Section III.B.1). The tunneling process is suppressed by the WKB factor<sup>5</sup>  $p \approx e^{-d\sqrt{2m\Delta}/\hbar}$ , where  $m$  is the effective mass of the tunneling entity. Transitions among substates are assumed to involve relatively local motions, such as the displacement of individual amino acid side chains,<sup>94</sup> for which  $m \approx 100 \text{ a.m.u.}$  Thus the WKB factor is  $p \approx 10^{-10}$ . Although this is a small number, we require that tunneling be slow on a chemical time scale, which is  $10^3 \text{ s}$  at the lowest temperature studied.<sup>27</sup> The tunneling rate is roughly the product of the WKB factor and an "attempt frequency" which we see must be less than  $10^7 \text{ s}^{-1}$ . Since the attempt frequency is approximately given by the vibrational frequency in one of the substates, this is untenable.

The X-ray data alone do not allow a decisive choice between the harmonic and multiple substate models. Taken together with the kinetic data, however, they cast considerable doubt on the viability of a picture in which protein dynamics consist of "hopping" among states at high temperature but a frozen distribution over states at low temperature. The magnitude of atomic motions in proteins, while perhaps larger than initially expected, is nonetheless so small that tunneling among substates would be significant at absolute zero. A corollary of this conclusion is that the observed motions at low temperatures are consistent with the magnitude of quantum zero-point motion, since it is this motion which would account for the tunneling process (cf. Section III.A.1). In fact, we shall see below that this claim, which strongly supports the applicability of the harmonic model, can be made quantitative.

The problem of correlation among atomic motions can be avoided in metalloproteins such as myoglobin through the use of Mössbauer spectroscopy, which is sensitive only to the single iron atom of the molecule. The Mössbauer spectrum of a simple nucleus, centered at energy  $E_0$  and of natural linewidth  $\Gamma$  is given by<sup>101</sup>

$$I(E) = \frac{\Gamma}{4} \int_{-\infty}^{\infty} d\tau e^{-i(E-E_0)-\frac{1}{2}\Gamma|\tau|} \langle e^{i\vec{k}\cdot\vec{x}(\tau)} e^{-i\vec{k}\cdot\vec{x}(0)} \rangle, \quad (\text{III.B.3.2})$$

where  $\vec{x}(t)$  is the position of the nucleus,  $\vec{k}$  is the wavevector of the Mössbauer radiation, and  $\hbar=1$ ; the spectrum is normalized such that if there is no motion of the nucleus  $I(E_0)=1$ . If the motion of the nucleus is dominated by a single normal mode (of frequency  $\omega$  and homogeneous linewidth  $\gamma$ ), then by analogy with the generating function in Eq. (III.A.3.39) we have

$$\langle e^{i\vec{k}\cdot\vec{x}(t)} e^{-i\vec{k}\cdot\vec{x}(0)} \rangle = e^{-S(2\bar{\nu}+1)} \exp[S(\bar{\nu}+1)e^{i\omega\tau-\gamma|\tau|} + S\bar{\nu}e^{-i\omega\tau-\gamma|\tau|}], \quad (\text{III.B.3.3})$$

where  $S = k^2 \langle (\delta x)_0^2 \rangle$ , with  $\langle (\delta x)_0^2 \rangle$  the mean-square magnitude of the zero-

point motion in the direction of  $\bar{k}$ , and  $\bar{\nu} = (e^{\hbar\omega/k_B T} - 1)^{-1}$  is the mean number of phonons. Substituting into Eq. (III.B.3.2), we find that  $I(E)$  has a number of components separated by multiples of the frequency  $\omega$ . The component centered on  $E_0$  is the true "recoilless" Mössbauer line:

$$I_0(E) = \frac{\Gamma}{4} e^{-S(2\bar{\nu}+1)} \sum_{n=0}^{\infty} \frac{[S^2 \bar{\nu}(\bar{\nu}+1)]^n}{(n!)^2} \frac{\Gamma + 4\gamma n}{(\frac{1}{2}\Gamma + 2\gamma n)^2 + (E - E_0)^2}. \quad (\text{III.B.3.4})$$

Integrating over  $E$ , we obtain the recoil-free fraction, or Lamb-Mössbauer factor,

$$f' = e^{-S(2\bar{\nu}+1)} I_0[2S\sqrt{\nu(\bar{\nu}+1)}], \quad (\text{III.B.3.5})$$

where  $I_0$  is a modified Bessel function. In the limit  $T \rightarrow 0$  this becomes

$$f'(T=0) = e^{-S}, \quad (\text{III.B.3.6})$$

so that the recoil-free fraction measures the zero-point motion; with many modes it may be shown that this result generalizes to<sup>93</sup>

$$-\ln f'(T=0) = k^2 \sum_{\mu} \langle (\delta x_{\mu})_0^2 \rangle, \quad (\text{III.B.3.7})$$

so that we can determine the total mean-square zero-point motion of the iron atom along the direction of  $\bar{k}$ .

Two problems arise in the measurement of zero-point motion by Mössbauer spectroscopy, at least in the case of myoglobin. First, the experiments were done on polycrystalline samples, so that rather than measuring  $f'$  the data give

$$f_{app} = \int d\phi f'(\bar{k}), \quad (\text{III.B.3.8})$$

where  $d\phi$  is an element of solid angle measuring the relative orientation of  $\bar{k}$  and some principal axis of the iron motion. This orientational averaging leads to an underestimate of the motion, since it includes orientations which are orthogonal to the

principal axis of the zero-point motion.

More serious is the fact that the determination of  $f'$  requires the measurement of an absolute absorption cross-section for the Mössbauer radiation, with all the attendant difficulties of absolute (as opposed to relative) measurements. Because of these difficulties, the series of papers which discuss the Mössbauer and crystallographic studies of myoglobin contain a most confusing discussion of the calibration problem.<sup>94,95,97,98</sup> Essentially each of the published reports advocates a different calibration standard, and the results from different standards are quite disparate. Suggestions range from the assumption that the zero-point motion of the iron atom vanishes, which begs the question at hand, to comparisons between the Mössbauer results and the crystallographic Debye-Waller factors, which have been discredited above.

These uncertainties suggest that the magnitude of the zero-point motion should be taken as being bounded from below by the most careful absolute measurements, and that calibration schemes which are external to the Mössbauer experiment should be disregarded. The data then determine the zero-point motion of the iron atom in myoglobin to be

$$\delta x_{z.p.}(Fe) = 0.0063 \text{ nm.} \quad (\text{III.B.3.9})$$

In terms of the model Hamiltonians presented here, the displacement of the iron atom may be written

$$\begin{aligned} x_{Fe} &= \sum_{\mu} A_{\mu}(Fe) q_{\mu} \\ &= \pi^{1/2} \sum_{\mu} A_{\mu}(Fe) \frac{(a_{\mu} + a_{\mu}^{\dagger})}{(2\omega_{\mu})^{1/2}}, \end{aligned} \quad (\text{III.B.3.10})$$

where  $a_{\mu}^{\dagger}$  and  $a_{\mu}$  create and annihilate phonons of the  $\mu^{\text{th}}$  mode, and the  $A_{\mu}(Fe)$  are a set of coefficients which transform from the normal mode coordinates  $q_{\mu}$  back to

atomic coordinates. The zero-point motion is therefore

$$\delta x_{z.p.}^2(Fe) = \frac{\hbar}{2} \sum_{\mu} A_{\mu}^2(Fe) \omega_{\mu}^{-1}. \quad (\text{III.B.3.11})$$

The coefficient  $A_{\mu}(Fe)$  may be determined from the effects of isotopic substitution of the iron on the frequency of the  $\mu^{\text{th}}$  mode.<sup>102</sup> This fact is central to the results which follow, and is therefore derived explicitly.

The momentum of the iron atom is given by

$$\begin{aligned} P_{Fe} &= m_{Fe} \sum_{\mu} A_{\mu}(Fe) \dot{q}_{\mu} = m_{Fe} \sum_{\mu} A_{\mu}(Fe) p_{\mu} \\ &= im_{Fe} \left(\frac{\hbar}{2}\right)^{1/2} \sum_{\mu} A_{\mu}(Fe) \omega_{\mu}^{1/2} (a_{\mu}^{\dagger} - a_{\mu}). \end{aligned} \quad (\text{III.B.3.12})$$

If we change the mass of the iron atom from  $m_{Fe}$  to  $m_{Fe} + \Delta m_{Fe}$ , we introduce a perturbation to the Hamiltonian

$$\begin{aligned} H_1 &= \frac{1}{2} \left( \frac{1}{m_{Fe} + \Delta m_{Fe}} - \frac{1}{m} \right) P_{Fe}^2 \\ &\approx -\frac{\hbar}{4} \Delta m_{Fe} \sum_{\mu\nu} A_{\mu}(Fe) A_{\nu}(Fe) (\omega_{\mu} \omega_{\nu})^{1/2} [a_{\mu}^{\dagger} a_{\nu} + a_{\mu} a_{\nu}^{\dagger} \\ &\quad - a_{\mu}^{\dagger} a_{\nu}^{\dagger} - a_{\mu} a_{\nu}]. \end{aligned} \quad (\text{III.B.3.13})$$

To lowest order in perturbation theory, only those terms with  $\mu = \nu$  contribute, and these give

$$\Delta H = -\frac{\hbar}{2} \Delta m_{Fe} \sum_{\mu} A_{\mu}^2(Fe) \omega_{\mu} (a_{\mu}^{\dagger} a_{\mu} + 1/2). \quad (\text{III.B.3.14})$$

This is equivalent to changes in frequency of each mode such that

$$\Delta\omega_{\mu} = -\frac{1}{2}\Delta m_{Fe} A_{\mu}^2(Fe)\omega_{\mu}, \quad (\text{III.B.3.15})$$

or

$$A_{\mu}^2(Fe) = -2\omega_{\mu}^{-1} \frac{\Delta\omega_{\mu}}{\Delta m_{Fe}}. \quad (\text{III.B.3.16})$$

The coefficients required to determine the contribution of each normal mode to the motion of the iron atom can thus be determined by vibrational spectroscopy and isotopic substitution, as promised. This approach is the opposite of the usual one in describing normal modes: rather than trying to determine the combinations of atomic motions which give rise to a particular normal mode, we are trying to find the combination of normal modes which gives rise to a particular atomic motion.

Table III-1 shows the frequencies and frequency shifts for those modes which are found, by Raman spectroscopy, to change frequency with isotopic substitution of the iron atom in myoglobin.<sup>78,103</sup> These few Raman-visible modes determine a zero-point motion of the iron of  $\delta x_{z,p}(Fe) = 0.0031 \text{ nm}$ , which is half that determined from the Mössbauer spectra. There are no arbitrary parameters in the conversion from the Raman data to atomic displacements.

Table III-1: Fe-sensitive modes in Myoglobin		
Frequency $\omega$	$\Delta\omega/\Delta m_{Fe}$	Reference
$501 \text{ cm}^{-1}$	$0.75 \text{ cm}^{-1}/a.m.u.$	103
$304 \text{ cm}^{-1}$	$0.5 \text{ cm}^{-1}/a.m.u.$	78
$220 \text{ cm}^{-1}$	$1.0 \text{ cm}^{-1}/a.m.u.$	78

If the harmonic model is correct, then the remainder of the iron zero-point motion is contributed by modes which are not resolved in the Raman spectra. It is unlikely that these modes are of very high frequency, since higher frequency modes must have very large frequency shifts ( $\approx 10 \text{ cm}^{-1}$ ) upon isotopic substitution in order to have a significant effect on the zero-point motion. Instead we focus on the low-

frequency region, where the Raman spectra display only a broad smear. This region contains the "breathing" modes of the protein, which are critical to the interpretation of the data discussed in the preceding sections, as well as the acoustic modes of the crystal.

The contribution from the acoustic modes of the crystal may be estimated from the Debye model,<sup>104</sup> which states that the zero-point motion of an element in a simple lattice (one molecule per unit cell) is

$$\delta x_{z.p.}^2 = \frac{9\hbar^2}{4Mk_B\Theta_D}, \quad (\text{III.B.3.17})$$

where  $M$  is the molecular mass and  $\Theta_D$  is the Debye temperature,

$$\Theta_D = \frac{\hbar s}{k_B} \left( \frac{6\pi^2}{v_c} \right)^{1/3}, \quad (\text{III.B.3.18})$$

with  $s$  the speed of sound in the crystal and  $v_c$  the volume of a unit cell. All of these parameters, except  $s$ , are known from the crystal structure of myoglobin<sup>79,95</sup> Since the crystal is largely made up of water, we take  $s \approx 10^3 \text{ m-s}^{-1}$  as a first estimate. This yields  $\Theta_D \approx 7.4 \text{ K}$  and  $\delta x_{z.p.} \approx 0.0029 \text{ nm}$ . More generally we have

$$\delta x_{z.p.} = (s/10^3 \text{ m-s}^{-1})^{-1/2} (0.0029 \text{ nm}), \quad (\text{III.B.3.19})$$

so that we must have  $s \geq 10^3 \text{ m-s}^{-1}$  to be consistent with the Mössbauer data. In addition, the Debye model predicts that for  $T \gg \Theta_D$ ,

$$\langle (\delta x)^2 \rangle = (\delta x_{z.p.})^2 \frac{4T}{\Theta_D}. \quad (\text{III.B.3.20})$$

With  $s = 10^3 \text{ m-s}^{-1}$ , the results for room temperature are  $\langle (\delta x)^2 \rangle \approx 1.4 \times 10^{-3} \text{ nm}^2$ , which is a bit large to be consistent with the overall Debye-Waller factor of myoglobin measured<sup>79</sup> crystallographically.<sup>105</sup> Thus we must

have  $s > 10^3 m-s^{-1}$  for consistency with both the Mössbauer and crystallographic data.

Previous discussions of motion in protein crystals have assumed that the contributions from the lattice as a whole are dominated by a static translational disorder, while the calculations presented here suggest that the dynamic disorder contributed by the acoustic phonons may be significant. The possibility of such a contribution--which is necessarily temperature dependent--means that "corrections" for lattice disorder may be misleading, and that apparent mean-square displacements should be reported without such corrections. In particular the acoustic modes contribute to the apparent motions observed in both the crystallographic and Mössbauer experiments, so that attempts to separate lattice contributions by comparing the two experimental results are not valid. Much more quantitative discussions could be given if the sound velocity in myoglobin crystals was measured directly, or if the acoustic modes could be detected by inelastic neutron scattering. For the present we can conclude only that the acoustic modes do not contribute all of the zero-point motion left after the Raman-visible modes are accounted for. We therefore turn to a consideration of the protein breathing modes.

The zero-point motion which we must account for has a magnitude of less than 0.003 nm. In the notation of Eq. (III.B.3.3), this corresponds to  $S = k^2 \langle (\delta x)^2 \rangle \approx 0.048$ , since for  $^{57}\text{Fe}$   $k = (2\pi/0.0086) \text{ nm}^{-1}$ . From Eq. (III.B.3.4) it may be shown that the broad line components resulting from a mode of frequency  $\omega$  become significant at a temperature  $T$  such that  $2Sk_B T \approx \hbar\omega$ . Experimentally,<sup>97,97,99</sup> the broad line components become significant near  $T \approx 200 \text{ K}$ , which determines  $\omega \approx 15 \text{ cm}^{-1}$ . Thus it seems that the zero-point motion of the iron atom observed by the Mössbauer effect may be interpreted in terms of a combination of vibrational modes:

- [1] The intermediate frequency ( $\approx 200\text{--}500\text{ cm}^{-1}$ ) modes, which most likely reflect vibrations of the bonds between the iron and its immediate neighbors. This contribution is roughly half the total.
- [2] The acoustic modes of the entire crystal, which are intermolecular motions. This contribution is probably significantly less than half the total.
- [3] A small number (perhaps one) of very low frequency breathing modes of the protein. The frequency of this mode,  $\approx 15\text{ cm}^{-1}$ , is very nearly that required to account for both the temperature-dependent reaction rate and the splittings observed in the infrared spectrum of the bound carbon monoxide, as discussed in the previous two sections.

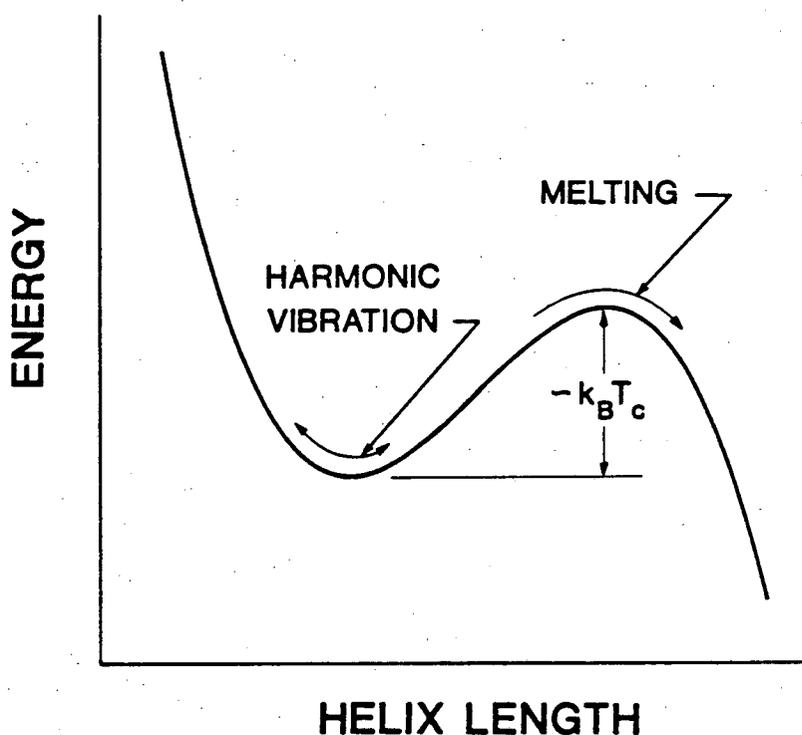
The low frequency mode is thus the essential link among the kinetic and spectroscopic data.

All the critical features of the Mössbauer spectra observed in myoglobin can be understood by further consideration of the nature of low frequency vibrational modes in proteins. To make the discussion concrete, let us imagine the lowest frequency stretching mode of an alpha helix; in fact myoglobin consists of eight helical segments, the longest of which is  $\approx 1.5\text{ nm}$  and by Eq. (III.A.2.15) will therefore have a vibrational frequency of  $\approx 15\text{ cm}^{-1}$ , which just what we require. The regular positions of the amino acids along the helix constitute a quasi-one dimensional lattice. Above some critical temperature  $T_c$  this lattice melts and the helix denatures into a random coil configuration. In the random coil the fluctuations in the end-to-end length of the helix, which is the normal coordinate of the stretching mode, become comparable to the dimensions of the helix itself, in the same way that the thermal fluctuations in the atomic positions in a crystal become comparable to the lattice spacing as the melting point is approached. A strictly harmonic model of the helix motion cannot account for this transition, but a simple modification shown in Fig. III-17 can. This picture is borrowed from the theory of melting in solids,<sup>106</sup> and the basic idea is that the stretching

mode is unstable against unraveling of the helix, but to "see" this instability requires some activation energy; this activation energy then determines the denaturation temperature. The central result of this theoretical approach is that, as the denaturation temperature  $T_c$  is approached, the apparent frequency of the vibrational motion decreases as

$$\omega(T) \approx \omega(1 - T/T_c)^{1/2}, \quad (\text{III.B.3.21})$$

so that at  $T_c$  the frequency goes to zero and the vibration becomes a quasi-free diffusion of the random coil.



**FIGURE III-17** Anharmonicity in collective vibrational motions of a protein. A model potential surface describing the melting of an alpha helix.

The variation of the effective frequency of motion with temperature is significant in many biological systems because macromolecular melting temperatures are not far above room temperature. Thus myoglobin denatures at  $T \approx 75\text{ C}$  or  $350\text{ K}$ , so that the effective frequency declines by a factor of two as temperature is raised from  $0\text{ K}$  to

250 K. With these parameters, it may be shown that the recoil-free fraction calculated from Eq. (III.B.3.15) will exhibit a precipitous decline above  $\approx 220$  K, in agreement with experiment.<sup>96-99</sup>

We can go further with this single-mode model and interpret the temperature dependence of the Mössbauer lineshape. The experimental spectra have been described as consisting of a temperature dependent mixture of two Lorentzian lines, each with temperature dependent widths, the broad line being approximately one hundred times the width of the narrow line.<sup>97,99</sup> Theoretically, it is clear from Eq. (III.B.3.4) that the spectra should be described by a more complex lineshape, although with experimental error two Lorentzians may be all that can be resolved. Whatever the lineshape, the broad components of the Mössbauer spectrum have a width  $\approx \gamma$ , the natural linewidth of the vibrational mode. The Mössbauer experiments therefore determine  $\gamma \approx 10^9$  s<sup>-1</sup>, which is in excellent agreement with the value of  $\gamma$  required to explain the non-exponential kinetics in the analysis of Sect. III.B.1 above..

If the series expansion of Eq. (III.B.3.4) is truncated at two terms, we obtain exactly a double Lorentzian lineshape. The narrow line has an integrated cross-section which falls monotonically with temperature, as observed, while the broader line first increases and then decreases, having a maximum cross-section at  $T \approx 200$  K, again in agreement with experiment. If we now include the effects of the further terms in the series, the apparent linewidths of the two Lorentzians will seem to broaden, since the successive terms of the series are increasingly broad; this broadening is also in agreement with observation. More quantitative fits to the data--which might resolve, for example, a temperature dependence of the natural linewidth  $\gamma$ --would be misleading because of the orientational averaging effect noted above. This qualitative agreement, however, reinforces the conclusion that these data *can* be understood in terms of simple model Hamiltonians, and that there is no evidence for the existence of conformational substates.

The central point regarding the interpretation of the Mössbauer data on the basis of a (nearly) harmonic model Hamiltonian is that the parameters of this Hamiltonian directly determine the observable quantities in independent experiments. In particular, a complete understanding of the Mössbauer experiments requires the existence of a low frequency mode of motion in the protein which has the same characteristics required by the analysis of *both* the kinetic data and the infrared spectra. Finally, the Mössbauer data provide independent evidence to demonstrate that the large coupling constant ( $S \approx 50$ ) for the breathing mode is in fact consistent with the very small atomic displacements ( $\approx 0.03 \text{ nm}$ ) observed upon ligand binding.

Physically, low frequency modes involve concerted motions of large numbers of atoms, so that a large motion of the normal mode requires only a small motion of any particular atom. The fit to the Mössbauer data, which shows that this mode contributes less than  $0.003 \text{ nm}$  to the zero-point motion of the iron atom, makes this argument quantitative. A coupling constant  $S$  corresponds to a displacement of  $S^{1/2}$  times the zero-point motion, so that the combined analyses of the kinetic and Mössbauer data predict that the iron atom is displaced by less than  $0.03 \text{ nm}$  in the transition from bound to photodissociated ligand. This is in excellent agreement with experiment,<sup>79</sup> and this agreement completes the argument that *all* of the structural and spectroscopic data are consistent with a single set of parameters, the same set of parameters required to account for the anomalous kinetic behavior of the molecule.

#### 4. Towards more detailed tests of the theory

The preceding sections have shown that the observation of non-exponential decay in the ligand binding reactions of the heme proteins may be interpreted as a quantum mechanical effect in the dynamics of biological molecules, and that the parameters required by this interpretation are confirmed by independent experiments. In principle I could present this argument in the opposite direction: the extant structural and spectroscopic data determine certain parameters in a model Hamiltonian description of

heme protein dynamics, and calculations of the functional behavior of the molecule--the kinetics of ligand binding--from these parameters are in agreement with experiment.

In fact, an *a priori* calculation of the kinetics from the extant spectroscopic data is not possible because the data are at present too sparse. Nonetheless, the ability to do such a calculation, in addition to its intrinsic interest, would provide the most detailed test of the theory presented here. In particular, the fact that the myoglobin-CO interaction exhibits such unusual temperature dependent non-exponential kinetics and anomalous isotope effects means that calculations of the kinetic behavior could be compared against much more than simply a single rate constant or activation energy.

Detailed tests of the theory will require a much larger body of data. In particular, attention should be paid to the following critical experiments:

- [1] High-resolution measurement of the CO stretching spectrum for *all* isotopic combinations of Fe, C, and O (twenty-seven in all).

The low-frequency mode is expected, from the fit to the Mössbauer data, to shift by  $\approx 0.1 \text{ cm}^{-1}$  upon isotopic substitution of the iron, and the spacing of the infrared lines should shift by a comparable amount. Together with the anomalous isotope effects predicted from anharmonic coupling, this suggests that the desired spectral resolution is  $\approx 0.01 \text{ cm}^{-1}$ .

- [2] Similarly precise measurements, either in infrared or Raman spectra, of the Fe-CO bond stretch and bending modes,<sup>75</sup> as well as the Fe-His mode.<sup>78</sup> In each case the published spectral lineshapes exhibit broadening and asymmetries which probably reflect unresolved anharmonic couplings.

One reason for the lack of resolution in the Raman technique may be the presence of multi-phonon effects due to the extreme low frequency mode,<sup>107</sup> but this effect may allow several properties of the low-frequency mode to be measured indirectly. Again extremely accurate measurements of isotopic frequency shifts are crucial because the

deviations of these shifts from their "naive" values (cf. Section III.B.2) provides a measure of coupling among low and high frequency modes.

- [3] Tests of the anharmonic model Hamiltonians by detection of vibrational overtones in the Raman spectrum.
- [4] Direct measures of anharmonic coupling by an "Infrared-Raman" method, in which the transition between ground and excited states of a high frequency mode is used to resonantly enhance inelastic scattering. The frequency shifts in the scattering then measure the frequencies of modes coupled to the high frequency mode.

This effect, illustrated in Fig. III-18, has not (to my knowledge) been detected in any system, although the analogous effect in vibration-rotation coupling is a familiar feature of gas-phase molecular spectroscopy. The suggestion that the number of modes coupled to the CO stretch is quite small makes this system ideal for such a measurement.

- [5] Careful comparisons between the atomic motions predicted from isotopic frequency shifts, in particular of the low-frequency modes detected by their effects in high-frequency infrared and Raman spectra, and those observed in Mössbauer spectra.

Ideally these spectra should be taken from aligned single crystals so that motions in different directions can be resolved and the difficulties of orientational averaging (cf. Section III.B.3) can be avoided. Similarly it would be attractive if the other spectroscopic data could be taken from crystalline specimens, so that there are no differences among samples used for different techniques.

These and other experiments, when interpreted along the lines introduced in this Chapter, will provide a complete characterization of the vibrational potential surface in each of the accessible electronic states. The preliminary estimates presented above strongly suggest that the more detailed information accessible in these further

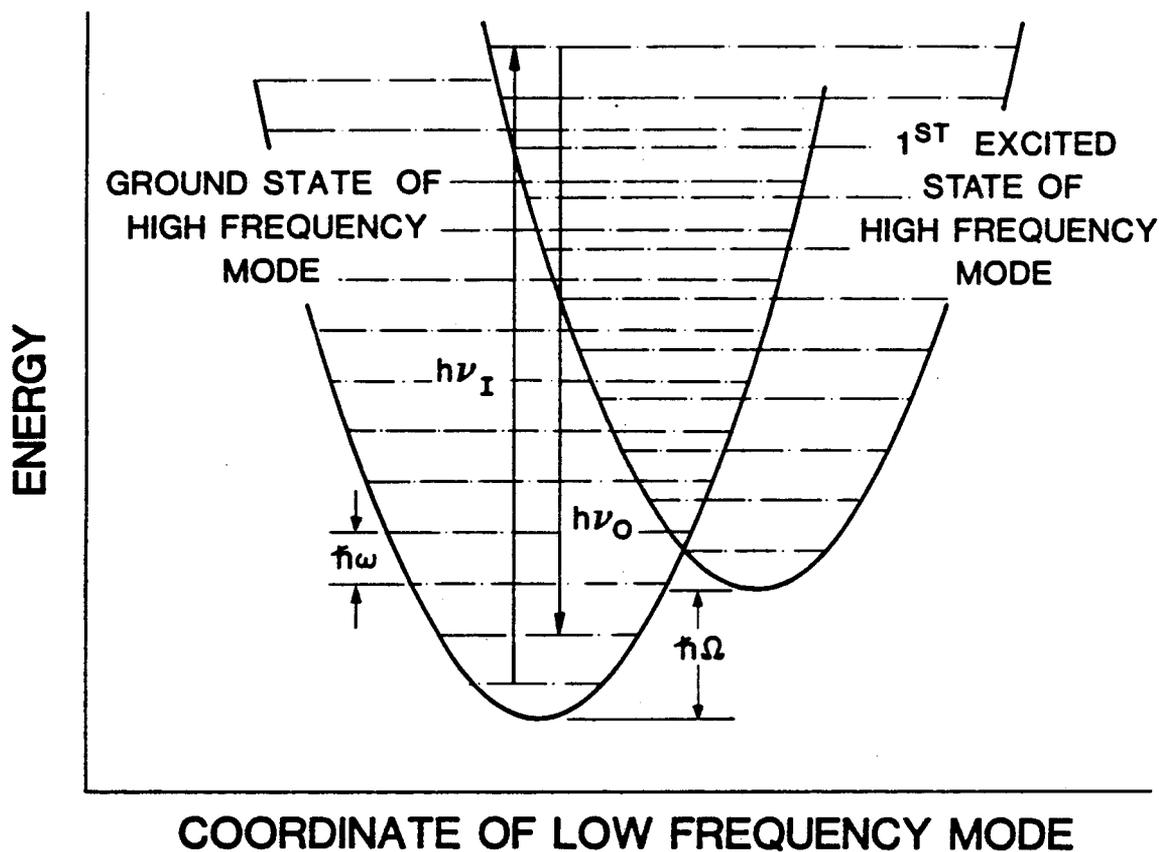


FIGURE III-18 The "Infrared-Raman effect."

experiments will allow successful calculations of reaction rates and their distributions as a function of temperature. This approach is *a priori* in that it involves a quantum mechanically rigorous calculation of the kinetic behavior from the dynamics of the molecule, but it is empirical in that no attempt is made to account for the origin of the vibrational spectrum in terms of individual inter-atomic interactions. This intermediate level of description, which is so naturally formalized in simple model Hamiltonians, shows great promise to be the next major step in the road toward a quantitative understanding of the physical basis for the molecular events in biological systems.

### C. Coherence in the picosecond regime

#### 1. When is an intermediate an intermediate?

The second major quantum mechanical effect predicted in Section III.A.4 is that of coherence for transitions which occur on a time scale comparable to vibrational relaxation. As noted in Section III.A.1, the primary events of photosynthesis certainly fall into this regime. The following sections therefore address the possibility that the predicted coherence may be detected experimentally.

The first step in any quantitative analysis of a system is to decide which states of the system are relevant. Thus in developing model Hamiltonians we must decide how many electronic states to include. It is attractive to identify the number of electronic states in the model Hamiltonian with the number of intermediate states observed in kinetic experiments, and essentially this is what has been done in previous work on biological systems.<sup>34</sup> There are, however, significant problems about what we mean by "intermediate" and "state" on a picosecond time scale.

Any spectroscopic experiment designed to follow a sequence of transitions makes the assumption that each state of the system has a unique spectrum. Thus if we plot the optical absorption as a function of time and frequency,<sup>108</sup>  $A(t; \Omega)$ , we should find that the spectrum consists of a superposition of spectra from the various states of the system, weighted by their populations:

$$A(t; \Omega) = \sum_i p_i(t) A_i(\Omega). \quad (\text{III.C.1.1})$$

The second conventional assumption is that the time course of the populations in each state can be written as a sum of exponentials, where the number of terms in the sum is one less than the number of states of the system:

$$p_i(t) = p_i(\infty) + \sum_{j=1}^{N-1} C_{ij} e^{-\lambda_j t}. \quad (\text{III.C.1.2})$$

These interpretive postulates seem innocuous, but in fact they are both in error once the transitions of interest begin to compete with vibrational relaxation.

To begin, we have seen in Section III.A.2 that the populations of electronic states are not expected to decay exponentially except in the slow regime. One way of understanding how and when exponential decay ceases to be a good approximation is to consider a highly averaged description of energy flow from electronic states to vibrational states and ultimately to the heat bath. If there were no vibrational relaxation, the energy of the electronic and vibrational degrees of freedom taken together would be conserved. Thus if we use the Pauli matrices to describe the two electronic states of the system [cf. Eq. (III.A.3.45)], and  $\nu$  is the number of phonons of frequency  $\omega$ , we have

$$\frac{d(\hbar\omega\nu + \epsilon \langle \sigma_z \rangle)}{dt} = 0, \quad (\text{III.C.1.3})$$

where  $\epsilon$  is the energy gap between the two states. Inclusion of vibrational relaxation at some rate  $\gamma$  to the equilibrium phonon population  $\bar{\nu}$  yields

$$\frac{d\nu}{dt} = -\gamma(\nu - \bar{\nu}) - \frac{\epsilon}{\hbar\omega} \frac{d\langle \sigma_z \rangle}{dt}. \quad (\text{III.C.1.4})$$

But the time dependence of the electronic populations are given by

$$\frac{d\langle \sigma_z \rangle}{dt} = -\frac{\langle \sigma_z \rangle - \sigma_z(\infty)}{T_1(\nu)}, \quad (\text{III.C.1.5})$$

where the dependence of the relaxation time  $T_1$  on the phonon population has been written explicitly.

Equations (III.C.1.4) and (III.C.1.5) describe the relaxation of the coupled electronic-vibrational system in the "effective temperature" approximation, in which it is assumed that the vibrational modes reach *some* temperature much more rapidly than they reach their equilibrium temperature; this approach is analogous to the spin temperature approximation in magnetic systems.<sup>109</sup> These equations are non-linear, and

do not have exponential solutions. To gain some insight into the solutions they do possess, we can go to the high-temperature limit, in which  $\nu = k_B T_{eff}/\hbar\omega$  and  $T_1^{-1} \approx e^{-E_a/k_B T_{eff}}$ . Using these expressions and expanding in  $\delta\nu = \nu - \bar{\nu}$ , one obtains

$$\frac{d\langle\sigma_z\rangle}{dt} = -\frac{1}{T_1}\left[1 + \frac{\hbar\omega}{k_B T} \frac{E_a}{k_B T} \delta\nu\right](\langle\sigma_z\rangle - \sigma_z(\infty)), \quad (\text{III.C.1.6a})$$

$$\begin{aligned} \delta\dot{\nu} = & -\gamma\left[1 - (\gamma T_1)^{-1}(\epsilon/k_B T)(E_a/k_B T)(\langle\sigma_z\rangle - \sigma_z(\infty))\right]\delta\nu \\ & + \frac{\epsilon}{\hbar\omega} \frac{1}{T_1}(\langle\sigma_z\rangle - \sigma_z(\infty)). \end{aligned} \quad (\text{III.C.1.6b})$$

We see that disaster--instability of the phonon populations--strikes as soon as  $\gamma T_1 < (\epsilon/k_B T)(E_a/k_B T)$ . In order for the "slow" approximation to be valid, we must therefore have

$$\gamma T_1 \gg \frac{\epsilon}{k_B T} \frac{E_a}{k_B T}, \quad (\text{III.C.1.7})$$

which is much more stringent than the naive condition  $\gamma T_1 \gg 1$ . Given that this condition is met, an approximate solution for the time dependence of the electronic populations is

$$\langle\sigma_z(t)\rangle = \sigma_z(\infty) + (\langle\sigma_z(0)\rangle - \sigma_z(\infty))e^{-t/T_1}e^{-x(1-e^{-t/T_1})}, \quad (\text{III.C.1.8})$$

and

$$x = \frac{1}{\gamma T_1} \frac{\epsilon}{k_B T} \frac{E_a}{k_B T} (\langle\sigma_z(0)\rangle - \sigma_z(\infty)).$$

This solution is shown in Fig. III-19 for various values of the parameter  $x$ , which measures the deviation from an exponential decay.

The approximation developed here corresponds to a simple physical picture. As the reaction proceeds, energy is transferred from the electronic to the vibrational

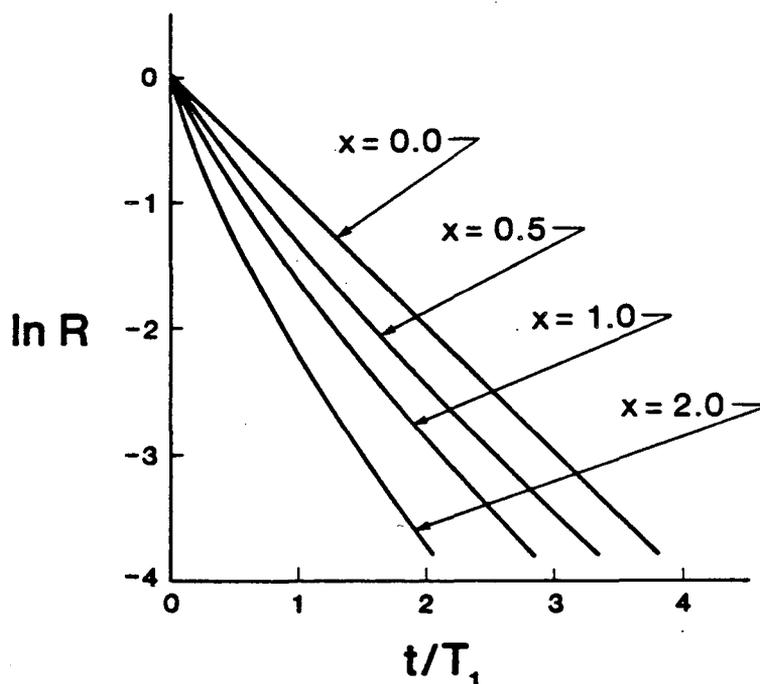


FIGURE III-19 Time course of picosecond electronic transitions--effective temperature approximation. The figure shows Eq. (III.C.1.8) for different values of the parameter  $x$  as indicated, with

$$R(t) = \frac{\langle \sigma_z(t) \rangle - \sigma_z(\infty)}{\langle \sigma_z(0) \rangle - \sigma_z(\infty)}$$

degrees of freedom. If the reaction becomes too fast, this energy cannot be completely dissipated to the heat bath, and the vibrations of the molecule heat up; as they heat up they change the reaction rate, leading to a reaction rate which varies in time--a non-exponential decay. As the reaction rate becomes still faster, this approximate view breaks down, and we must make use of the Green's function techniques developed previously in Section III.A.3.

The second problem in dealing with picosecond events is that the spectrum associated with occupancy of a single electronic state may itself be time-dependent. Even if the radiative interactions are independent of the vibrational state of the system, and therefore seem to probe only the electronic states, the strong electron-phonon coupling in the Hamiltonian means that the details of the absorption spectrum will depend on both electronic and vibrational states. As applied to optical absorption, this is just a

restatement of the fact that the optical lineshape is affected by the vibrational modes. Although this fact is well known, two of its corollaries are not generally appreciated.

To begin, the electron-phonon coupling means that the optical absorption spectrum depends on the distribution over phonon states. On the time scale of vibrational relaxation, this distribution is not stationary, so that a given electronic state cannot give rise to a definite absorption spectrum. In addition, the excitation pulse used to initiate the experiment will leave behind a non-equilibrium distribution over the phonon states, and this perturbation to the distribution will be decaying on the same time scale as the transitions we are trying to observe. These effects give rise to further complications in the time dependence of the absorption measured at a given frequency, so that the apparent number of exponentials describing this time dependence can never be taken as the number of electronic states.

The second corollary of the vibrational-dependence of optical lineshape is that in a strongly coupled system there are *no* spectroscopic probes which are sensitive to electronic states alone. Perhaps the most surprising example of this is provided by magnetic resonance. Electron paramagnetic resonance spectroscopy makes use of radiation in the microwave region, where the energies are small and only the spin states should be involved. While this is true of the radiative interaction itself, it is not true of the radiationless interactions which give rise to the parameters describing the spin dynamics.

The primary charge separation reaction in photosynthesis proceeds from a singlet excited state of the donor, so that one electron of a spin-paired set is removed and transferred to the acceptor. As a result, the reaction produces *two* unpaired electrons, and the spins of these electrons can interact with each other as well as with nuclear spins at the donor and acceptor sites. These interactions can be probed by looking at the polarization of the electron spins,<sup>110</sup> or by detecting the new states which can be reached once the spin system has evolved into the triplet state.<sup>111</sup>

The electrons localized on the photosynthetic donor and acceptor, like any electrons localized on distinct sites, may exhibit a magnetic interaction mediated by virtual electron transfer; this interaction is termed superexchange. Anderson's<sup>112</sup> original theory of superexchange described *d* or *f* electrons in poor conductors, such as the transition metal oxides. Recently, however, this theory has been applied<sup>36,113,114</sup> to the case of photosynthesis, where we have seen that strong electron-phonon coupling is essential to our understanding of the electron transfer process. As discussed above, we suspect that any spectroscopic probe of such strongly coupled systems will be sensitive to both electronic *and* vibrational spectra. To illustrate this point I give here an analysis of superexchange in a simple two-site model Hamiltonian, but with electron-phonon coupling included according to the methods developed in Section III.A.3. The results show that electron-phonon couplings which promote rapid electron transfer quench the superexchange interaction. This result is opposite to the correlation between poor conduction and weak exchange found in the absence of vibronic coupling, and provides a natural explanation of some of the experimental observations on photosynthetic systems, as will be seen below.

Consider four states<sup>115</sup> of a donor/acceptor pair:

$$\begin{aligned} |(DA)^S\rangle &= |1\rangle & |(D^+A^-)^S\rangle &= |2\rangle \\ |(DA)^T\rangle &= |3\rangle & |(D^+A^-)^T\rangle &= |4\rangle \end{aligned}$$

where *S* and *T* denote singlet and triplet spin states, respectively. If the matrix element *V* for electron transfer ( $DA \rightarrow D^+A^-$ ) is spin independent, the general Hamiltonian is

$$\begin{aligned} \mathbf{H} = & H_1(\bar{p}, \bar{x}) c_1^\dagger c_1 + H_2(\bar{p}, \bar{x}) c_2^\dagger c_2 + V(c_1^\dagger c_2 + c_2^\dagger c_1) \\ & + H_3(\bar{p}, \bar{x}) c_3^\dagger c_3 + H_4(\bar{p}, \bar{x}) c_4^\dagger c_4 + V(c_3^\dagger c_4 + c_4^\dagger c_3), \end{aligned}$$

(III.C.1.9)

where  $(\vec{p}, \vec{x})$  denotes the momenta and positions of the nuclei, respectively. By analogy with the discussion of Sections III.A.3, we can construct state-dependent phonon operators and then transform to phonon operators which depend on the expectation values of the electron operators; when all this is done we obtain [compare Eq. (III.A.3.3)]

$$\begin{aligned}
 U\mathbf{H}U^\dagger = & \mathbf{H}_{ph} + \sum_i (\epsilon_i + D_i(\langle a_\mu, a_\mu^\dagger \rangle)) c_i^\dagger c_i - \sum_i D_i(\langle a_\mu, a_\mu^\dagger \rangle) \langle c_i^\dagger c_i \rangle \\
 & + V[F_{12}c_1^\dagger c_2 + F_{12}^\dagger c_2^\dagger c_1 + F_{34}c_3^\dagger c_4 + F_{34}^\dagger c_4^\dagger c_3]. \quad (\text{III.C.1.10})
 \end{aligned}$$

Again it is convenient to simplify to cases where the normal mode structure and frequencies do not depend on the electronic states, so that  $D_i = 0$ .

The magnitude of the superexchange matrix element is the difference in energy between states  $|2\rangle$  and  $|4\rangle$  which arises from the perturbation  $V$ . If we apply standard time-independent perturbation theory, then all of the vibrational sub-states of the electronic states must be enumerated and all of the matrix elements of  $F_{ij}$  (which are Franck-Condon factors) must be evaluated. To avoid these complexities, transform the Hamiltonian again, using the unitary operator  $S = e^{iW}$ , with  $W$  chosen such that  $\mathbf{H}_{int} + i[\mathbf{H}_0, W] = 0$ ;  $\mathbf{H}_{int}$  denotes all the terms of Eq. (III.C.1.10) which depend on  $V$ , and  $\mathbf{H}_0$  denotes all the  $V$ -independent terms. If we expand the transformed Hamiltonian to second order in  $V$ , terms linear in  $V$  do not appear and there is an added term,  $\Delta\mathbf{H} = (i/2)[\mathbf{H}_{int}, W]$ ; this term is the operator equivalent of the energy shifts resulting from second-order time-independent perturbation theory.<sup>101,116</sup>  $W$  is determined explicitly by noting that, in the interaction representation,  $dW/dt = -\mathbf{H}_{int}$ , so that

$$\Delta\mathbf{H}(t) = -\frac{i}{2} \int_{-\infty}^t d\tau [\mathbf{H}_{int}(t), \mathbf{H}_{int}(\tau)]$$

$$= -\frac{i}{2} \int_0^{\infty} d\tau [\mathbf{H}_{int}(t), \mathbf{H}_{int}(t-\tau)]. \quad (\text{III.C.1.11})$$

Evaluating this operator in states  $|2\rangle$  and  $|4\rangle$ , the superexchange element becomes

$$J_{eff}(t) = -\frac{V^2}{2} \text{Im} \int_0^{\infty} d\tau [e^{i(\epsilon_1 - \epsilon_2)\tau} F_{12}(t) F_{12}^\dagger(t-\tau) - e^{i(\epsilon_3 - \epsilon_4)\tau} F_{34}(t) F_{34}^\dagger(t-\tau)]. \quad (\text{III.C.1.12})$$

The superexchange matrix element is an operator--because of its dependence on the phonon creation and annihilation operators--and hence, with the vibrations held in thermal equilibrium, exhibits fluctuations in time. Even if there is no average superexchange interaction, there can be a finite spectral density to the fluctuations, which provides a new mechanism of spin dephasing in vibronically coupled systems.

The result of Eq. (III.C.1.12) may be compared with the expressions, given in Appendix D, for the transition probabilities  $W_{ij}$  among the states  $i$  and  $j$  in the slow regime:

$$W_{ij}(\epsilon_i - \epsilon_j) = \frac{V^2}{2} \text{Re} \int_0^{\infty} d\tau e^{i(\epsilon_i - \epsilon_j)\tau} \langle F_{ij}(t) F_{ij}^\dagger(t-\tau) \rangle. \quad (\text{III.C.1.13})$$

The average energy shifts calculated from Eq. (III.C.1.12) and the decay rates in Eq. (III.C.1.13) are, respectively, the real and imaginary parts of the self-energy functions, which are analytic functions of the energy gaps  $\epsilon_i - \epsilon_j$ . Thus there exists a dispersion relation

$$\langle J_{eff} \rangle = -P \int_{-\infty}^{\infty} \frac{d\xi}{\pi} \left\{ \frac{\Gamma_{12}(\xi)}{\epsilon_1 - \epsilon_2 - \xi} - \frac{\Gamma_{34}(\xi)}{\epsilon_3 - \epsilon_4 - \xi} \right\}. \quad (\text{III.C.1.14})$$

In certain approximations, discussed in Appendix D, the decay rates exhibit Gaussian

dependences on the energy gap (the "energy-gap law"),<sup>117</sup> while if only a few phonon modes are coupled to the transition one obtains a sum of Gaussians. In either case the calculation of the energy shifts reduces to evaluating the Hilbert transform of a Gaussian, or the plasma dispersion function.<sup>118</sup> For energy gaps at the peaks of the Gaussians, the energy shifts vanish, and hence the superexchange interaction is quenched.

For the special case of a single harmonic phonon mode with no frequency shifts and an inhomogeneous linewidth  $\gamma$ , explicit results are:

$$W_{ij} = V^2 \sum_{n=-\infty}^{\infty} \rho_n(S_{ij}; T) e^{-(\epsilon_i - \epsilon_j - n\omega)^2 / 2\sigma_{ij}^2}, \quad (\text{III.C.1.15})$$

$$\Delta E_{ij} = \sqrt{(2/\pi)\sigma_{ij}^2} \sum_{n=-\infty}^{\infty} \rho_n \Delta E_{ij}(n); \quad (\text{III.C.1.16})$$

$$\Delta E_{ij}(n) = -\frac{V^2(\epsilon_i - \epsilon_j - n\omega)}{\sigma_{ij}^2} \left[ 1 - \frac{(\epsilon_i - \epsilon_j - n\omega)^2}{3\sigma_{ij}^2} \right]$$

$$|\epsilon_i - \epsilon_j - n\omega| \ll \sigma_{ij}, \quad (\text{III.C.1.17a})$$

$$\Delta E_{ij}(n) = \frac{V^2}{\epsilon_i - \epsilon_j - n\omega} \left[ 1 + \frac{\sigma_{ij}^2}{2(\epsilon_i - \epsilon_j - n\omega)^2} \right]$$

$$|\epsilon_i - \epsilon_j - n\omega| \gg \sigma_{ij}, \quad (\text{III.C.1.17b})$$

where

$$\rho_n = \frac{4e^{-S_{ij}(2\bar{\nu}+1)}}{\gamma} \left( \frac{\bar{\nu}+1}{\bar{\nu}} \right)^{n/2} \frac{\hbar\omega(\pi/2)^{1/2}}{\epsilon_i - \epsilon_j} I_n [2S_{ij}z\sqrt{\bar{\nu}(\bar{\nu}+1)}], \quad (\text{III.C.1.18})$$

$$S_{ij} = \eta_{ij}^2 \quad \bar{\nu} = \left[ e^{\hbar\omega/k_B T} - 1 \right]^{-1} \quad \sigma_{ij} = 2\gamma/S_{ij}, \quad (\text{III.C.1.19})$$

and  $\langle J_{eff} \rangle = \Delta E_{34} - \Delta E_{12}$ ; the  $W_{ij}$  are taken from Appendix D. The quenching of the superexchange interaction near the rate maxima  $\epsilon_i - \epsilon_j = n\omega$  is apparent. In the limit of no vibronic coupling,  $S_{ij} \rightarrow 0$ , the sum in Eq. (III.C.1.16) reduces to a single term, and this term reproduces Anderson's<sup>112</sup> result.

In the statistical limit of many harmonic modes with incommensurate frequencies,

$$W_{ij} = V^2 \sqrt{\frac{8\pi}{s_{ij}}} e^{-(\epsilon_i - \epsilon_j - M_{ij})^2 / 2s_{ij}}, \quad (\text{III.C.1.20})$$

$$\Delta E_{ij} = -4 \frac{V^2 (\epsilon_i - \epsilon_j - M_{ij})}{s_{ij}} \left[ 1 - \frac{(\epsilon_i - \epsilon_j - M_{ij})^2}{3s_{ij}} \right]$$

$$|\epsilon_i - \epsilon_j - M_{ij}| \ll s_{ij}^{1/2}, \quad (\text{III.C.1.21a})$$

$$\Delta E_{ij} = 4 \frac{V^2}{\epsilon_i - \epsilon_j - M_{ij}} \left[ 1 + \frac{s_{ij}}{2(\epsilon_i - \epsilon_j - M_{ij})^2} \right]$$

$$|\epsilon_i - \epsilon_j - M_{ij}| \gg s_{ij}^{1/2}, \quad (\text{III.C.1.21b})$$

where

$$M_{ij} = \sum_{\mu} S_{ij}(\mu) \omega_{\mu} \quad s_{ij} = \sum_{\mu} S_{ij}(\mu) \omega_{\mu}^2 (2\bar{\nu}_{\mu} + 1) \quad (\text{III.C.1.22})$$

and the  $S_{ij}(\mu)$ , etc., are the natural generalizations of the single mode definitions. The result in Eq. (III.C.1.21b) is essentially that obtained by using<sup>113,114</sup> Anderson's theory and replacing the energy gap by the "vertical" energy gap in the vibronically coupled system; all other cases seem to be qualitatively different from what is seen in the absence of coupling.

A number of experimental studies<sup>110,111,119-121</sup> of the "primary" donor/acceptor pair in reaction centers from photosynthetic bacteria ( $P^+$  and  $I^-$  in the notation of Fig. III-6) suggest that  $\langle J_{eff} \rangle$  is very small, that the spin dephasing rates exceed those required by lifetime broadening, and that the spin interactions in the charge separated state are anisotropic. The small value of the exchange matrix element, when interpreted in terms of the Anderson<sup>112</sup> theory, is inconsistent<sup>114</sup> with the large values of the transfer matrix element required to account for the observed<sup>37,122</sup> picosecond charge separation. To resolve this discrepancy a multi-site model was proposed<sup>114</sup>; hopping among these sites then accounts for the excess spin dephasing. From the results presented here, there is no need to invoke the more complex model: the average superexchange element may be small because the energy gaps are adjusted to provide maximum rates for the charge separation reactions, which is plausible, and the spin dephasing can be generated by fluctuations in the exchange element due to vibronic coupling. Furthermore, the spin Hamiltonian of the intramolecular triplet state  $|3\rangle$  is anisotropic<sup>110</sup> so that the energy gap  $\epsilon_3 - \epsilon_4$  is orientation-dependent in a magnetic field; through the strong energy gap dependence of the superexchange interaction near the quenching points this can lead to an apparent anisotropy in the charge separated states  $|2\rangle$  and  $|4\rangle$ .

What are we to conclude from all this? Essentially no spectroscopic evidence for the existence of intermediate states on a picosecond time scale can be taken at face value. Those states which can be trapped and studied on longer time scales certainly do exist, but the properties of these trapped states--such as equilibrium free energies and absorption spectra--may be irrelevant to the quantitative analysis of picosecond events. On the positive side, these results imply, as in the example of the superexchange interaction, that the complete model Hamiltonian required for understanding the system may be much simpler than indicated at first, although it will be necessary to treat this Hamiltonian rigorously.

The major results from the analysis of dynamics in the picosecond regime are also suggestive of further experiments. We have seen that *no* process which competes with vibrational relaxation can be rigorously exponential in its time course, and future picosecond studies should be aimed at detecting this departure from exponential behavior. Equation (III.C.1.8) provides an alternative time course which can be compared with experiment for those transitions which are not competing too strongly with vibrational relaxation, and more general forms may be derived by the Green's function method presented above.

Another approach to the problems of the picosecond regime is to search for an effect which is diagnostic of all those quantum mechanical features which make the conventional experiments so difficult to interpret. This is the subject of the following section.

## 2. Long pulse vs. short pulse experiments

If the photosynthetic system, or any biological system, operates in a regime where quantum coherence effects are significant, we must take care to interpret all observations strictly in terms of quantum measurement theory. In particular, systems which display quantum coherence may exhibit different dynamics depending on the time scale of the measurement, and the following arguments use this fact to predict a feature of the optical spectroscopy of the photosynthetic reaction center which is diagnostic of the coherence effects; this prediction is confirmed by experiment.

If a system makes a transition from one state to another *coherently*, then on the time scale of the transition we have a quantum mechanical superposition of the two states. Thus the state vector at time  $t$  may be written

$$|\psi(t)\rangle = a_1|1\rangle + a_2|2\rangle. \quad (\text{III.C.2.1})$$

If we observe the system on this time scale, we collapse the wave function onto one of the two states and destroy the coherence. On the other hand, if we have two nearly

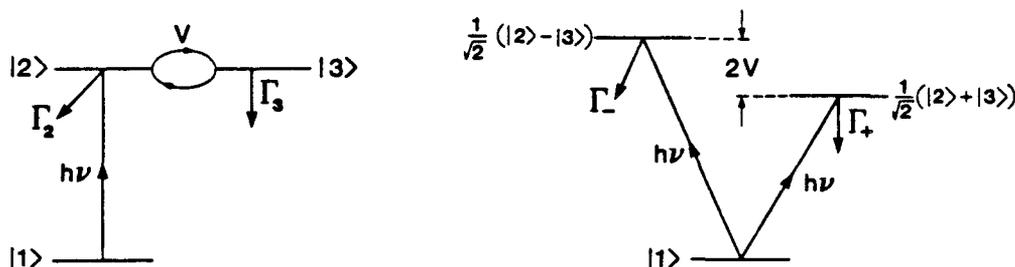
degenerate states coupled by a matrix element  $V$ , then on a time scale  $t \gg \hbar/V$  we can observe only two states (the symmetric and anti-symmetric combinations of  $|1\rangle$  and  $|2\rangle$ ) separated in energy by  $2V$ .

These considerations can be amplified to a three state model, in which the transition  $0 \rightarrow 1$  occurs by photon absorption and the mixing  $1 \rightarrow 2$  occurs *via* a matrix element  $V$  as before; this is illustrated in Fig. III-20. If we use an optical pulse short compared to  $\hbar/V$ , we can see the population of the state  $|1\rangle$  mix with that of state  $|2\rangle$ . If the pulse is long, and hence of well defined energy, we can in principle resolve the energy difference  $2V$  between the states  $|+\rangle$  and  $|-\rangle$  defined by

$$|\pm\rangle = \frac{1}{\sqrt{2}}[|1\rangle \pm |2\rangle]. \quad (\text{III.C.2.2})$$

On this scale the absorption process therefore carries the system from  $|0\rangle$  to either  $|+\rangle$  or  $|-\rangle$  depending on the frequency of the radiation. In particular, if we use white light<sup>123</sup> so that different frequencies have no definite phase relations, the symmetric and anti-symmetric states are populated with no definite phase relations: long pulse experiments give rise to an incoherent mixture of  $|+\rangle$  and  $|-\rangle$ , whereas short pulse experiments lead to a particular coherent superposition of these two states, namely state  $|1\rangle$ .

Although all the dynamics of radiative interactions on short and long time scales can be described within the Green's function formalism developed in Section III.A.3 and Appendix E, these basic ideas about coherence are all that is necessary for the analysis that follows. The critical point is that long pulse excitation takes the system into both states  $|1\rangle$  and  $|2\rangle$ , even though the absorption matrix element only connects the ground state with state  $|1\rangle$ . This may be understood by second order perturbation theory. We have a dipole matrix element  $D$  for the transition  $0 \rightarrow 1$  and a radiationless matrix element  $V$  for the transition  $1 \rightarrow 2$ . The radiative cross-section for the overall process  $0 \rightarrow 2$  is therefore



**FIGURE III-20** Coherence in a three-state system. Two states are coupled by a radiative matrix element, and one of these states is coupled by a radiationless matrix element to a third. On the left is the picture obtained from experiments on this system using short pulses of light, which *temporally* resolve the radiationless transition. On the right is the picture appropriate to long-pulse experiments, which in principle can *spectrally* resolve the splitting due to the radiationless matrix element.

$$\sigma_{0 \rightarrow 2} \approx \left| D \frac{V}{\Delta E} \right|^2, \quad (\text{III.C.2.3})$$

where the energy denominator  $\Delta E \approx \Gamma$ , the decay rate of the excited state, when the states are nearly degenerate. Thus the cross-section for reaching state  $|2\rangle$  is significant compared to that for reaching state  $|1\rangle$ ,

$$\sigma_{0 \rightarrow 1} \approx |D|^2, \quad (\text{III.C.2.4})$$

whenever the radiationless matrix elements are comparable to the decay rates of the states, which is precisely the condition for the coherent behavior described above.

To connect these ideas with the photosynthetic reaction center, we identify the three states with the ground state  $PI$ , the excited state of the donor  $P^*I$ , and the charge separated state  $P^+I^-$ , as shown in Fig. III-21. The treatment given above predicts that if the radiationless transition  $P^*I \rightarrow P^+I^-$  competes with vibrational relaxation, which it does (cf. Section III.A.1), then a long excitation pulse must directly populate the charge separated state, even though the dipole matrix element does not reach this state. Thus under normal light conditions the photosynthetic system may

proceed through the excited state of the primary donor only "virtually," and a portion of the optical absorption which drives photosynthetic charge separation does so directly without any intermediate in the chemical sense.

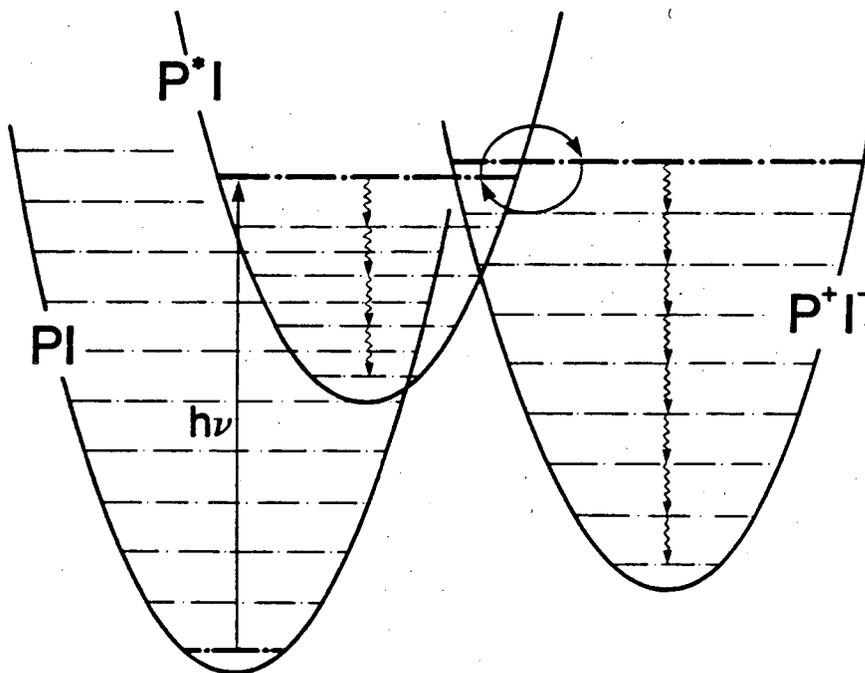


FIGURE III-21 Coherence in the primary events of photosynthesis. This figure is analogous to the left hand part of Fig. III-20--it shows how a particular photon frequency chooses three *vibronic* states of the reaction center to which the analysis of the text can be applied.

The well known absorption band of the primary photosynthetic process may thus be viewed as having two components of roughly equal magnitude (at least in perturbation theory), one of which corresponds to a direct transition to the charge separated state. The component which directly populates the state  $P^+I^-$  is associated with a large change in dipole moment, since it involves motion of an electron over several angstroms, an *intermolecular* rather than an *intramolecular* distance. Thus we expect that this absorption band will have an anomalously large Stark shift. In fact an anomalous Stark shift is, according to the present theory, the signature of any photo-induced electron transfer reaction which competes with vibrational relaxation.

Experiments on the isolated photosynthetic reaction centers of the bacterium *Rhodospseudomonas sphaeroides* demonstrate that the Stark effect of the absorption in the primary donor is nearly an order of magnitude larger than that of other pigmented components of the reaction center.<sup>124</sup> Thus experiment supports the central prediction of the theory. It is important to emphasize that the prediction of a large Stark effect is based strictly on the notion of quantum coherence in the primary charge separation step.

### 3. Testing the theory

By working out the theory of the anomalous Stark effect in more detail, it may be shown that the apparent dipole moment associated with the optical absorption is not constant across the full width of the absorption band. The theory thus predicts that the magnitude of the Stark shift will not be simply related to the absorption spectrum, as would be the case for a simple two-level system. More detailed observations of the Stark effect should be aimed at resolving these variations.

The Stark effect can also form the basis of a much more ambitious test of the present theory. This is because the Stark effect provides an optical probe of all three states involved in the primary charge separation, rather than just two as would be true if coherent mixing did not occur. Thus if we can extract some parameters of the three states and compare them with independent experiments, we can put the hypothesis of coherent mixing to a stringent test. In particular, we would like to determine how the vibrational degrees of freedom are coupled to the changes in electronic state  $PI \rightarrow P^*I$  and  $PI \rightarrow P^+I^-$ , since these coupling constants directly influence the rate of the charge separation reaction.

The changes in vibrational coordinate associated with an optical transition between two electronic states determine the intensities of the lines in the resonance Raman spectrum associated with that transition.<sup>125</sup> The Raman scattering process is a two-photon event, in which the system absorbs a photon of frequency  $\Omega$ , reaching an

excited electronic/vibrational state, and then emits a photon of frequency  $\Omega'$  by returning to the electronic ground state but a different vibrational level than the one initially populated. Thus, as illustrated in Fig. III-22, the frequency shift  $\Omega - \Omega'$  must be an integer multiple of the vibrational frequencies of the molecule, while the intensity of the inelastic scattering is related to the overlap of the vibrational states in the ground and excited electronic states, or Franck-Condon factors. Some of the theoretical issues involved in a rigorous calculation of Raman spectra in strongly coupled electron-phonon systems are discussed in Appendix E; the basic point, however, is that the Raman spectrum allows the measurement of the vibrational frequencies, linewidths and coupling constants of those modes which are coupled to the electronic transition. As a test of this determination, the parameters obtained from the Raman data can be used to reconstruct the detailed shape of the absorption spectrum.

In the photosynthetic system we have two electronic transitions superposed, namely  $PI \rightarrow P^*I$  and  $PI \rightarrow P^+I^-$ , but we can apply an electric field and use the Stark effect to shift the resulting absorption bands relative to one another. In the same way, the application of an electric field will shift the excitation profiles of the two components in the Raman spectrum. In this way we can measure the coordinate changes and other electron-phonon couplings which are relevant in the transitions among all three of the states involved in the primary photosynthetic reaction.

The "Raman Stark effect" proposed here, together with the interpretation of the magnetic interactions of the primary radical pair (cf. Section III.C.1), allows a complete determination of the parameters in the model Hamiltonians which have been used here to describe biomolecular dynamics. This determination can be made entirely by spectroscopic techniques, and no three-dimensional structure is required to understand the results at this level. The extraction of these parameters makes possible, through the theory presented in Sections III.A, a completely rigorous calculation of the time dependence of the electronic populations which will be observed in a

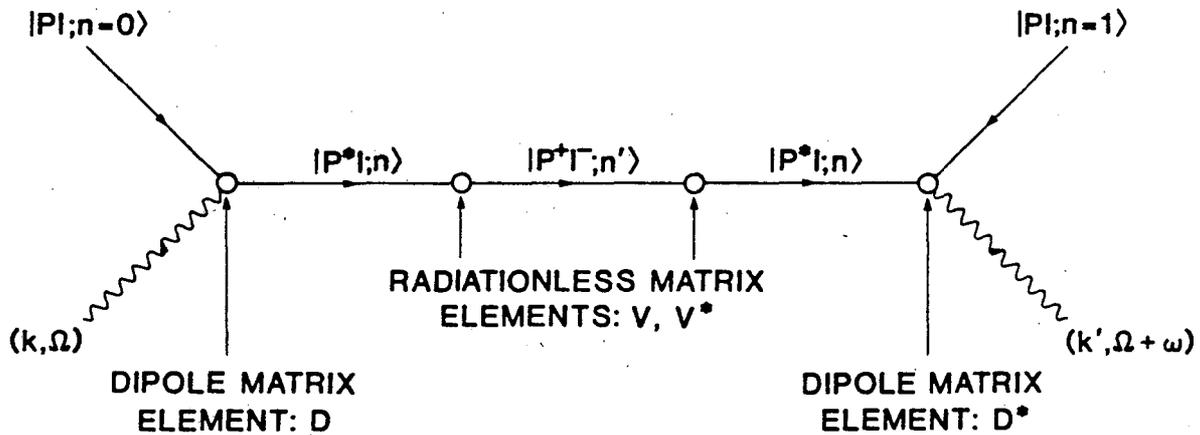


FIGURE III-22 Raman scattering from the primary photosynthetic absorption band. Following the analysis of Fig. III-21, the Raman process may include a virtual transition to the charge separated state, as shown in this Feynman diagram. Because the energy levels of the charge separated state are strongly dependent on the strength of an applied electric field, this part of the Raman cross-section will exhibit a large Stark shift, allowing it to be distinguished from more conventional Raman processes. In this way the vibronic coupling parameters (electron-phonon coupling constants) of both the excited ( $P^*I$ ) and charge separated ( $P^*I^-$ ) states can be measured from the Raman intensities and resonance excitation profiles.

picosecond experiment. As in the case of myoglobin (cf. Section III.B.4), such a calculation allows the most rigorous possible test of the theory presented here.

#### 4. Some conclusions

Sections III.B and III.C have been a search for two quantum effects in the dynamics of biological molecules: resonances in the slow regime and coherence in the picosecond regime. The experimentally testable predictions from these two effects are non-exponential decay and an anomalous Stark effect in photo-induced charge transfer. Both these predictions are confirmed by experiments, in myoglobin and photosynthetic reaction centers respectively. In the case of myoglobin independent experiments demonstrate that the interpretation of the observations in terms of quantum effects is plausible. In this sense the search has been an unqualified success: at least

two systems display behavior which is consistently interpreted in terms of non-trivial quantum mechanical contributions to the dynamics of biological molecules.

The arguments of this Chapter also demonstrate that the application of simple model Hamiltonians to biological molecules is helpful in the design of further experiments. In both of the systems considered, prospects appear quite good for a complete determination of the parameters of the model Hamiltonian using existing experimental techniques. Theory would then allow the direct calculation of biologically relevant quantities, namely the rates and quantum efficiencies of the reactions in which the molecules participate. No other theoretical approach has promised such a quantitative relation between the dynamics of biological molecules and their functional behavior, particularly in the ultra-fast time domain which characterizes the primary events of photosynthesis. Further, existing theories have had difficulty defining the experimental parameters which are critical to the understanding of biomolecular dynamics, as witnessed by the discussion at a recent symposium.<sup>126</sup>

Both of the systems discussed here display behavior which cannot be interpreted in classical terms without introducing great complexity. Thus the non-exponential decay in myoglobin may be described by assuming a distribution of molecular structures and assigning each structure a different reaction rate,<sup>100</sup> but this approach makes no prediction regarding the relations among the distribution of reaction rates and the distribution of vibrational frequencies observed by infrared spectroscopy. Nor does the structural model, which is an outgrowth of the structure-function hypothesis outlined at the beginning of this thesis, speak to the essential question of how the structure controls the rate. The approach taken here answers both these objections, albeit in quantum mechanical terms. Further, the present theory allows the identification of a new factor--the vibrational spectrum--in the array of molecular parameters which the organism can manipulate in its attempt to control its chemistry.

Thus the theoretical machinery required to understand the quantum effects which occur in biological molecules is also useful for examining the mechanisms by which the biological system accomplishes its task of regulating the course of chemical events. Another task of the organism, which was explored in Chapter II, is that of making measurements on its environment. Remarkably, the organism accomplishes this task with a level of performance which approaches the limits imposed by quantum mechanics. This ability is evidence of macroscopic quantum behavior, and this behavior can only be understood in terms of the microscopic quantum effects which occur at the level of individual biological molecules. Now that we have learned something about these effects, and developed a general theoretical approach to their description, we can return to the sensory systems. This is done in the following sections.

#### **D. Mechanisms for macroscopic quantum behavior in biological systems**

##### *1. Phonon super-radiance*

We have seen that a sufficiently rapid transition between electronic states will result in a "heating" of the molecular vibrations and ultimately in an instability of these vibrational modes. This may be understood as follows: the electronic transition transfers energy to the vibrations, while the rate of the electronic transition increases with temperature or equivalently vibrational energy. As a result the "heating" of the phonons by a fast transition makes the transition still faster, and instability results.

While this effect may be described classically such an approach obscures the fundamental quantum mechanical character of the phonon instability. The reason that the rate of an electronic transition depends on temperature is that the phonon emission which must accompany the transition is stimulated by those phonons which are thermally excited, while the transition can still occur at zero temperature because of spontaneous phonon emission (cf. Section III.A.1). Thus the instability is similar to that which occurs in a laser: the electronic transition results in phonon emission and stimulated phonon emission causes the transition to accelerate as it transfers more energy to

the phonons. The major difference from the stimulated emission of photons is that the electronic transition considered here is always multi-quantum; in contrast the coupling of electrons to photons contains primarily single-photon terms, except in unusual circumstances.

Let us imagine not a single molecule but rather a polymer, or one-dimensional lattice of molecules each with two electronic states, ground and excited. If all the molecules of the lattice are initially in their excited electronic states, the phonons emitted by one molecule can propagate to another site on the lattice, stimulating the electronic transition and the resulting phonon emission. This process resembles the super-radiant<sup>127</sup> emission of photons by a collection of excited atoms, and may be termed phonon super-radiance. Such an effect--because of its multi-quantum nature--has a number of interesting features which may be compared to ordinary super-radiance. What is of interest here, however, is that super-radiance is a mechanism for generating a coherent phonon population throughout the entire polymer, or at least some macroscopic length of it. If it can occur within the model Hamiltonians which describe the dynamics of biological molecules, then phonon super-radiance provides a natural means of explaining the macroscopic quantum effects which we require in order to understand quantum-limited detection in the sensory systems. This and the following sections have the goal of making this connection between macroscopic and microscopic quantum behavior plausible.

Again, the simplest model Hamiltonian can be used to demonstrate the phenomenon of phonon super-radiance. The lattice sites along the polymer are labeled  $n$  and each site has two electronic states, created by  $c_{n1}^\dagger$  and  $c_{n2}^\dagger$  and one phonon mode with creation operator  $a_n^\dagger$ . The Hamiltonian of each individual lattice site is then [compare Eq. (III.A.3.40)]

$$H_n = \epsilon(c_{n2}^\dagger c_{n2} - c_{n1}^\dagger c_{n1}) + \omega a_n^\dagger a_n$$

$$+ V(F_n c_{n1}^\dagger c_{n2} + F_n^\dagger c_{n2}^\dagger c_{n1}) + \text{damping}, \quad (\text{III.D.1.1})$$

where

$$F_n = e^{-1/2\eta^2} \cdot e^{\eta(a_n - a_n^\dagger)}. \quad (\text{III.D.1.2})$$

Coupling among adjacent molecules in the polymer creates a phonon band so that

$$a_n = \frac{1}{\sqrt{N}} \sum_k e^{ikn} a_k, \quad (\text{III.D.1.3})$$

and

$$\sum_n \omega a_n^\dagger a_n = \sum_k \omega(k) a_k^\dagger a_k. \quad (\text{III.D.1.4})$$

The full Hamiltonian is thus

$$\mathbf{H} = \mathbf{H}_0 + V e^{-\eta^2} \sum_n [c_{n1}^\dagger c_{n2} \cdot \exp\{\eta' \sum_k (e^{+ikn} a_k - e^{-ikn} a_k^\dagger)\}] + H.C.], \quad (\text{III.D.1.5})$$

where  $\eta' = N^{-1/2}\eta$ , and

$$\mathbf{H}_0 = \epsilon \sum_n (c_{n2}^\dagger c_{n2} - c_{n1}^\dagger c_{n1}) + \sum_k \omega(k) a_k^\dagger a_k + \text{damping}. \quad (\text{III.D.1.6})$$

The phenomenon of phonon super-radiance can be studied in several ways. One is the transient method, in which we try to solve the equations of motion for the coupled electron and phonon operators starting from some initial condition where the electronic populations are "inverted." Another method is to explicitly couple the electronic subsystem to a "pump" which replenishes the excited state and maintains the inversion; it is this steady-state version of the problem which we expect to be relevant to amplification mechanisms in the sensory systems. Both these approaches involve

the solution of an explicitly non-equilibrium problem, with all of the attendant difficulties.

The essential features of phonon superradiance can in fact be understood by means of a zero-temperature, equilibrium Green's function analysis. To begin, we note that the energy of the electronic excitation is generally much larger than thermal energies, so that if significant "heating" of the vibrational modes occurs then essentially all of the phonon energy is non-thermal in origin; this allows us to consider the limiting case of zero temperature. For field-theoretic perturbation theory, however, the only special feature of the zero-temperature case is that the ground state is non-degenerate, so that the Gell-Mann and Low theorem is valid and may be used to prove cancellation of disconnected diagrams and other useful points.<sup>48,56</sup> All of these properties are shared by the "pseudo-ground state" corresponding to all molecules in the upper of their two electronic states. This pseudo-ground state is a zero-temperature state provided that the temperature  $T \rightarrow 0^-$  rather than the usual  $T \rightarrow 0^+$ ; for a discussion of negative absolute temperatures in two-level systems see Ref. 109. Thus we can study the interaction of a phonon system with an inverted electronic population by the same  $T=0$  methods developed for the normal case in Section III.A.3; the only difference is that in Eq. (III.A.3.32) the electronic state populations  $n_i$  must be chosen appropriately.

Using these results we can proceed directly. The quantities of interest are the phonon Green's functions in the interacting system, which are obtained from the (un-normalized) generating functional analogous to that for the electronic Green's functions in Eq. (III.A.3.22):

$$\Lambda'[J] = \langle T[\exp\{-i \int dt H_{int}(t) - i \sum_k \int dt (J_k^\dagger(t) a_k(t) + J_k(t) a_k^\dagger(t))\}] \rangle. \quad (\text{III.D.1.7})$$

The electron operators which appear in  $H_{int}$  may be replaced by dummy variables, as in Eq. (III.A.3.25), and a similar trick may be used for the phonon operators [cf. Eq. (III.A.3.38)]. For the interaction Hamiltonian of Eq. (III.D.1.5), all of the integrals over electronic variables are Gaussian and may be done exactly. The remaining integrals may be expanded in powers of the matrix element  $V$  and the terms of the expansion rearranged to form the phonon self-energy. This procedure is straightforward if tedious, and when all the dust clears, the phonon self-energy is independent of  $k$  and is given in the time domain by

$$\begin{aligned} \Sigma(\tau) = & -i4V^2\eta^2 e^{-\eta^2\theta(\tau)} e^{i\epsilon\tau} \\ & \times \exp\left[\frac{\eta^2}{N} \sum_k e^{-i\omega(k)\tau} e^{-\gamma_k|\tau|}\right], \end{aligned} \quad (\text{III.D.1.8})$$

where  $\gamma_k$  is the damping constant of the  $k^{\text{th}}$  mode in the absence of coupling to the electronic states. To simplify matters, I assume that the phonons form a narrow band, so that  $\omega(k) \approx \omega$  and  $\gamma_k \approx \gamma$ , and further that  $\epsilon = n\omega$ . Then the transformation to the frequency domain is straightforward, and

$$\Sigma(\Omega) = -\frac{in(n-1)^{-1}}{1 + i\frac{(\Omega - \omega)}{(n-1)\gamma}} 4\frac{V^2}{\gamma} e^{-S} \frac{S^n}{n!}, \quad (\text{III.D.1.9})$$

where  $S = \eta^2$ . Comparing this result with the zero-order Green's function

$$D = \frac{1}{\Omega - \omega - i\gamma}, \quad (\text{III.D.1.10})$$

it is clear that the sign of the self-energy corresponds to a negative damping constant, so that for sufficiently large self-energy (rapid transitions among the electronic states) instability will result, as promised.

The instability of the phonons, as measured by the fact that the poles of the Green's function cross the real axis, is analogous to other instabilities in many-body

systems. An example of this is provided by superconductivity,<sup>56</sup> in which the attractive interaction between electrons leads to an unstable pole in the two-particle Green's function. This indicates that the initial choice of the ground state is actually an unstable state, and this error must be remedied by self-consistent determination of the new ground state and the Green's functions which describe excitations above this state.

For the electronic instability of superconductors one must introduce new Green's functions  $\langle c(t)c(t') \rangle$  and  $\langle c^\dagger(t)c^\dagger(t') \rangle$  in addition to the usual  $\langle c^\dagger(t)c(t') \rangle$ ; the magnitude of the anomalous Green's function then becomes an order parameter in the sense of Landau theory of phase transitions. In the phonon instability the anomalous Green's functions are similarly  $\langle a(t)a(t') \rangle$  and  $\langle a^\dagger(t)a^\dagger(t') \rangle$ . The fact that these Green's functions are non-zero means that the phonons, rather than being in the ground state or even an excited state of definite phonon number, are in a coherent state<sup>128</sup>  $|\alpha\rangle$  such that  $a|\alpha\rangle = \alpha|\alpha\rangle$ . The amplitude of the coherent state now becomes the order parameter of the system, being zero for conditions of stability and non-zero once instability takes over. The fact that  $\alpha$  has a definite phase while the effective Hamiltonian is symmetric under the transformation  $\alpha \rightarrow e^{i\phi}\alpha$  implies that the state reached through the super-radiant instability is a broken-symmetry state, also as in the case of the superconductor. These analogies between the instabilities of a polymer and those of the electron gas in a superconductor are essentially the same as those considered by Haken<sup>45</sup> for the case of a laser and its transition from incoherent to coherent emission as the pump power is increased.

These arguments demonstrate that, if each subunit catalyzes a chemical reaction of sufficient rate, the phonon band can exhibit a super-radiant instability, and through this instability some phonon modes acquire a coherent oscillation amplitude throughout the polymer. The condition for the super-radiant instability is roughly that noted in Eq. (III.C.1.7), namely a vibrational relaxation rate  $\gamma \ll R\epsilon E_a/(k_B T)^2$ ,

where  $R$  is the rate of the "pump" reaction,  $E_a$  is its activation energy, and  $\epsilon$  is the energy gap for this reaction. Typical energy-yielding reactions in biology, such as the hydrolysis of adenosine triphosphate (ATP) have<sup>129</sup>  $\epsilon \approx 0.5 \text{ eV}$ , and individual steps in the reaction pathway will have  $E_a \approx 1.0 \text{ eV}$ , which means that near room temperature the rate of the reaction changes by a factor of three for a ten degree increase in the temperature.<sup>130</sup> Experimental estimates of phonon lifetimes in biological polymers, obtained from the widths of spectral lines observed in Brillouin scattering,<sup>41</sup> are  $\gamma \approx 10^9 \text{ s}^{-1}$ , which requires a reaction rate of  $R \geq 1.2 \times 10^6 \text{ s}^{-1}$  if phonon super-radiance is to occur. Reaction rates on the microsecond time scale are observed for enzyme-catalyzed and energy-converting reactions in biology,<sup>13</sup> so that the occurrence of super-radiance, however surprising, is actually quite plausible.

## 2. Building the perfect amplifier

Can the macroscopic coherence which results from the super-radiant instability help us to understand the approach to quantum-limited noise performance in sensory amplification? In particular, we would like to address the problem in the inner ear, where the most rigorous argument for quantum-limited detection has been made (cf. Sections II.C.2-4). In this case, the signals which must be amplified are mechanical in origin and range in frequency from  $\approx 10 \text{ Hz}$  to  $\approx 10^5 \text{ Hz}$  in bats and porpoises.<sup>131</sup> Clearly the phonons participating in the super-radiant instability must have frequencies much higher than these signal frequencies, so some sort of parametric amplification is indicated.

To see how the "high" (on the audible scale; low on the molecular scale) frequency phonons in a polymer can couple to mechanical signals, consider the phonon Hamiltonian in the presence of a strain  $x_n$  between the subunits at sites  $n$  and  $n + 1$ :

$$H_{ph} = \omega \sum_n a_n^\dagger a_n + \sum_n \left( A + \frac{\partial A}{\partial x} x_n \right) [a_n^\dagger a_{n+1} + a_{n+1}^\dagger a_n] + \text{damping},$$

where  $A$  is the coupling amplitude between adjacent subunits. Passing to  $k$  states as above,

$$\mathbf{H}_{ph} = \mathbf{H}_{ph}^{(0)} + \frac{\partial A}{\partial x} \sum_{k,k'} a_k^\dagger a_{k'} (e^{+ik'} + e^{-ik}) \sum_n e^{i(k-k')n} x_n. \quad (\text{III.D.2.2})$$

If we consider some normal mode of polymer motion with generalized coordinate  $q$ , then  $x_n = f_n q$ , so that the Hamiltonian which determines the interaction between the phonons and the macroscopic motion is

$$\mathbf{H}_{int} = \frac{\partial A}{\partial x} \frac{1}{\sqrt{N}} \sum_{k,k'} f(k'-k) (e^{+ik'} + e^{-ik}) a_k^\dagger a_{k'} q, \quad (\text{III.D.2.3})$$

where

$$f(k) = \frac{1}{\sqrt{N}} \sum_n e^{-ikn} f_n. \quad (\text{III.D.2.4})$$

This result may be rewritten more simply as

$$\mathbf{H}_{int} = \sum_{k,k'} g(k;k') a_k^\dagger a_{k'} q. \quad (\text{III.D.2.5})$$

But since  $q = q_0(b^\dagger + b)$ , where  $q_0$  is the zero-point motion and  $b^\dagger$  creates the vibrational quanta of the macroscopic motion, we obtain

$$\mathbf{H}_{int} = q_0 \sum_{k,k'} g(k;k') a_k^\dagger a_{k'} (b^\dagger + b). \quad (\text{III.D.2.6})$$

This is precisely the Hamiltonian of a parametric amplifier. The "para-amp" is one of the few devices known to achieve the quantum limit to amplifier performance, provided that one of the modes is driven to a large amplitude by an external source.<sup>132,133</sup> Thus, to the extent that phonon super-radiance drives at least one phonon mode into a large amplitude coherent state, the polymer will behave as a perfect amplifier, with the macroscopic motion as the input mode and the internal phonon

modes as the output modes. As shown above, such coherent oscillation is indeed the limiting case of the super-radiant instability.

Consider that some phonon mode  $k_0$  acquires a coherent amplitude  $\alpha_0(t) = \alpha_0 e^{-i\omega_0 t}$  as a result of the super-radiant instability. If this amplitude is sufficiently large, the Hamiltonian can be approximated as

$$\mathbf{H}_{int}(t) \approx q_0 \sum_k [\alpha_0 g(k; k_0) a_k^\dagger e^{-i\omega_0 t} + \alpha_0^* g(k_0; k) a_k e^{+i\omega_0 t}] (b^\dagger + b), \quad (\text{III.D.2.7})$$

which describes a device in which motion of the polymer at frequency  $\omega$  results in the generation of phonons at frequencies  $\omega_0 \pm \omega$  in the polymer: parametric up-conversion. In addition, these high-frequency phonons will act back on the macroscopic motion, applying an apparent force  $F_{app} \approx q_0 g(k; k_0) a_k^\dagger \alpha_0 e^{-i\omega_0 t}$ . The effect of this feedback will be to change the response of the polymer to external forces, as described in Section II.B.3. In the simplified model of Eq. (III.D.2.7), this change in response may be calculated straightforwardly because the interaction Hamiltonian is quadratic in boson operators; recall that  $\alpha_0$  is so large that the  $k_0$  mode behaves classically. The result is that

$$\frac{\tilde{q}(\Omega)}{\tilde{F}(\Omega)} = 2q_0^2 \frac{\Omega_0 + \Sigma(\Omega)}{(\Omega_0 - \Omega - i\Gamma)(\Omega + \Omega_0 + i\Gamma) - 2\Omega_0 \Sigma(\Omega)}, \quad (\text{III.D.2.8})$$

where  $\Omega_0$  and  $\Gamma$  are the resonance frequency and damping constant of the macroscopic motion in the absence of feedback, while the self-energy term is

$$\Sigma(\Omega) = |q_0 \alpha_0|^2 \sum_k \frac{g(k; k_0) g(k_0; k) [\omega_0 - \omega(k)]}{[\omega_0 - \omega(k)]^2 - [\Omega + i\gamma_k]^2}, \quad (\text{III.D.2.9})$$

where  $\gamma_k$  is the *effective* damping constant of the  $k^{\text{th}}$  phonon mode in the presence of the super-radiant instability.

As expected from the description of parametric up-conversion, the self-energy has a pole whenever the signal frequency is close to being a difference between two phonon frequencies. This function thus has very different behavior depending on whether or not the spacing between adjacent modes--which is the inverse of the time required to from one end of the polymer to the other--is comparable to the signal frequency. This point may be seen by writing

$$\omega(k) \approx \omega_0 + \frac{\partial \omega}{\partial k} (k - k_0), \quad (\text{III.D.2.10})$$

where, since<sup>101</sup>  $k = \pi n/L$  ( $L$  is the length of the lattice and  $n$  is an integer smaller than the number of lattice points),

$$\omega(k) \approx \omega_0 + \frac{\pi}{L} \frac{\partial \omega}{\partial k} \Delta n. \quad (\text{III.D.2.11})$$

But  $\partial \omega / \partial k = c$ , the phonon propagation velocity, so that

$$\omega(k) = \omega_0 + \frac{\pi}{\tau} \Delta n, \quad (\text{III.D.2.12})$$

where  $\tau = L/c$  is the propagation time. The self-energy therefore becomes

$$\Sigma(\Omega) \approx |q_0 \alpha_0|^2 \sum_n \frac{g_n^2 n \pi \tau^{-1}}{n^2 - (\tau/\pi)^2 (\Omega + i\gamma_n)^2}. \quad (\text{III.D.2.13})$$

In this expression the resonant structure is apparent.

If the resonant structure of the self-energy is observable, then the length of the polymer determines the resonant frequency of the detector. This is an attractive hypothesis, particularly in the case of the inner ear where we have seen (cf. Section II.C.3) that no passive mechanical features of the detector can provide a resonance. Since typical signal frequencies are one kilohertz, while typical polymer lengths are one micron, such a resonance effect is significant if the phonon velocity is  $\approx 10^{-3} \text{ m-s}^{-1}$ ,

which seems quite slow. Acoustic modes in a linear polymer have  $c = (Y/\rho)^{1/2}$  where  $Y$  is the Young's modulus and  $\rho$  is the density; for actin and other proteins this implies (cf. Section II.C.2 for parameters estimates)  $c \approx 4 \times 10^3 \text{ m-s}^{-1}$ . On the other hand, the optical phonons of the polymer have very small propagation velocities, so that the dispersion relations appear as in Fig. III-23. If the vibrational modes of the individual subunits are in the  $10\text{--}50 \text{ cm}^{-1}$  range then the acoustic and optical modes have dispersion relations that cross. Near this crossing the two modes will mix, and the resulting two modes can have arbitrarily small velocities. The dynamics of biological molecules are apparently such as to cause the resonant character of the parametric amplification and feedback to be significant.

The arguments presented here demonstrate that the phenomenon of phonon super-radiance in polymers is a viable mechanism for generating macroscopic coherence in a biological system, and further that such coherence forms the basis for a perfect amplifier together with its feedback to the mechanics of polymer movement. This perfect amplifier with feedback is exactly what we found to be necessary in order to understand the detection of sub-angstrom displacements by the stereocilia in the inner ear. It is therefore tempting to identify the quasi-crystalline actin polymers which make up the stereocilium<sup>135,136</sup> and its supporting structures within the receptor cell<sup>137,138</sup> as the polymers which participate in the transition to super-radiance.

As discussed in Section III.C.3, the stereocilium is not naturally resonant in the audible range. Indeed typical signal frequencies  $\Omega$  are much less than either  $\Omega_0$  or  $\Gamma$ , so that if Eq. (III.D.2.8) describes the stereocilium it may be approximated as

$$\frac{\bar{q}(\Omega)}{\bar{F}(\Omega)} = 2q_0^2 \frac{\Omega_0 + \Sigma(\Omega)}{\Omega_1^2 - 2\Omega_0\Sigma(\Omega)}, \quad (\text{III.D.2.14})$$

where  $\Omega_1^2 = \Omega_0^2 + \Gamma^2$ . With Eq. (III.D.2.13) for the self-energy, this expression not only determines a single resonance frequency, it determines a whole family of possible

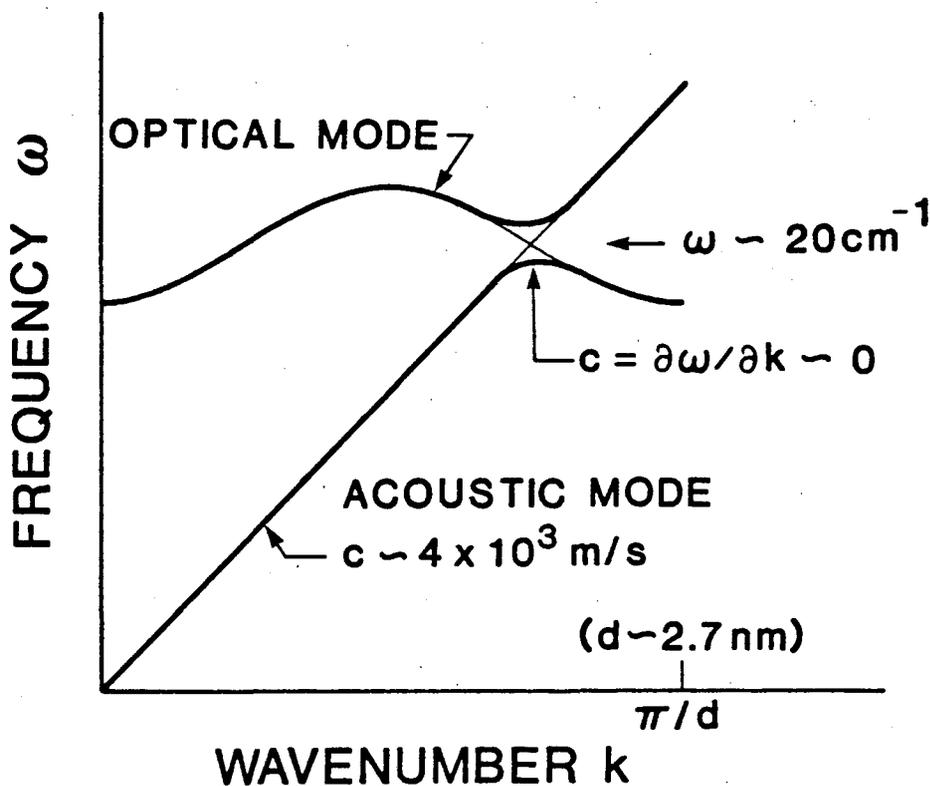


FIGURE III-23 Phonon dispersion in a biological polymer. The parameters given are chosen to be approximately those expected for the actin filaments of the stereocilia in the inner ear. The lattice spacing is known from diffraction measurements,<sup>134</sup> while the helical structure of the lattice implies that the dominant interactions are between second nearest neighbors along the lattice; this will give rise to the peaked structure in the optical mode dispersion. Note that at relatively high wavenumber the optical and acoustic modes cross with opposite slopes in their dispersion relations, so that the coupled modes have short wavelengths but very small group velocities as indicated.

resonance frequencies given by the poles of the response function:

$$\Omega_n = \frac{\pi}{\tau} \left[ n^2 - n \frac{2|g_n q_0 \alpha_0|^2 \pi \Omega_0}{\Omega_1 \tau} \right]^{1/2} - i\gamma_n. \quad (\text{III.D.2.15})$$

Of all these resonances, one will in general dominate the others, leading to a response function as shown in Fig. III-24. The index  $n$  of the dominant resonance depends on the magnitude of  $\alpha_0$ , which is controlled by the strength of the pump driving the

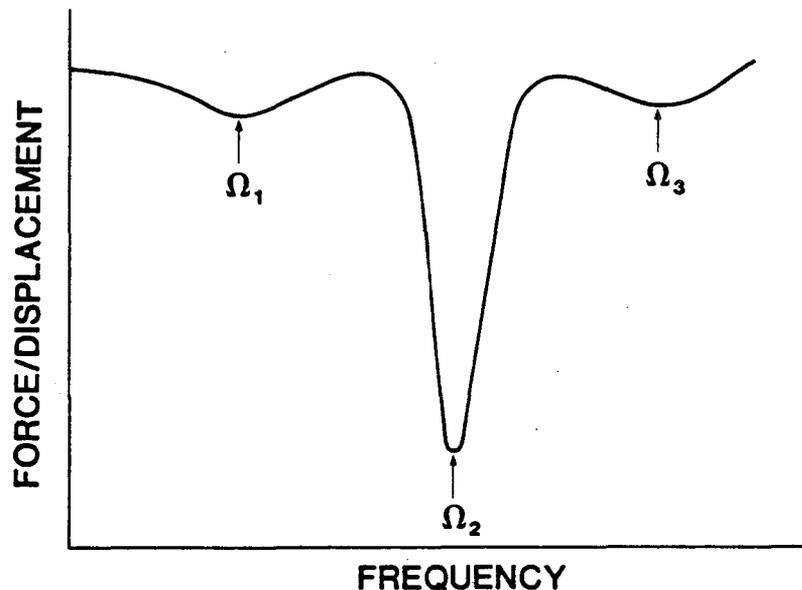


FIGURE III-24 Multiple resonances predicted in the active filters of the inner ear stereocilia--a schematic rendering of Eq. (III.D.2.14).

super-radiance, and on the variation of  $\gamma_n$  with  $n$ , which also depends on the characteristics of the pump. In addition, the resonance frequency can be shifted slightly by variations in  $g_n$ , which is related to the distribution of strain through the polymer and hence to the packing geometry of the polymer array. The critical point, however, is that the possible resonance frequencies of the detector are determined by the length of the polymer,  $\Omega_n \approx n\pi c/L$ .

### 3. Amplifier dynamics and experiments in the inner ear

The analysis of phonon super-radiance, together with the resulting amplification and mechanical feedback, leads us finally to a consistent physical picture of quantum-limited detection in the inner ear:

- [1] Enzymatic reactions within the stereocilium drive a phonon instability in the polymers which comprise that structure; this instability results in a macroscopic coherence of the phonon amplitude throughout the polymer.

- [2] Coupling between macroscopic motion of the stereocilium and internal phonon modes leads to parametric amplification of the mechanical signal and mechanical feedback. This amplification is quantum-mechanically "perfect," achieving the limiting noise performance allowed by the uncertainty principle.
- [3] The feedback synthesizes a narrow-band mechanical response for the stereocilium, the resonant frequency being controlled by the length of the polymers and the chemistry of the pump.
- [4] The active filter removes the thermal noise of the stereocilium, amplifies the signal so that is above the transducer noise (cf. Section II.C.4), and is limited only by the quantum noise in the amplifier, as described in Section II.B.3.

Each of these phenomena *can* occur within reasonable models for the dynamics of biological molecules, as has been shown in the previous two sections. In addition, something like these phenomena *must* occur if we are to understand quantum-limited detection; this was the major conclusion of Chapter II. The question which remains is whether there is any direct experimental test of these ideas.

In principle, the phonon instability could be directly observed by examining the Stokes-to-anti-Stokes ratio of Raman scattering from the polymers. This ratio measures<sup>139</sup> an average phonon "temperature" which will be higher than the ambient temperature if phonon super-radiance occurs. In practice, Raman spectra of biological samples are very weak unless they are resonantly enhanced, and there is no obvious chromophore in the actin polymer. It has recently proven possible, however, to record Raman spectra from "uncolored" proteins using the resonant enhancement which results from the ultraviolet absorption bands of aromatic amino acids.<sup>140</sup> If such a "Raman label" can be found in the stereocilium, then the high effective concentration of polymer should allow for measurement of the phonon temperature, although this will be by no means an easy experiment.

More accessible tests of the present theory are to be found in detailed measurements of the frequency response of the receptor cell. In particular, the peculiar characteristics of the feedback from the internal phonons to the macroscopic motion results in a pattern of multiple resonances, all of which are related to the length of the polymer. These effects may be understood qualitatively by imagining that local strain of the stereocilium injects (or depletes) phonons which then propagate up and down the polymer. When these phonons return to the point which was strained, they produce a feedback force, but this feedback is delayed by the propagation time. The system thus mimics feedback through a delay-line, which of course produces a sequence of resonances at frequencies such that the delay (propagation time) is an integer multiple of the signal period.

In fact, a number of experimental studies of frequency responses in inner ear organs display fine-structural features which may be interpreted as the secondary resonances which result from finite-delay feedback. These studies, some of which are illustrated in Fig. III-25, include both the responses of single cells to pure tones<sup>141,142</sup> and the frequency-dependence of multi-tone interactions,<sup>141,143</sup> and have been done on species ranging from frogs to humans (using acoustic and psychophysical techniques). These experiments are, however, only suggestive, and more detailed measurements will be required to define the precise relations among primary and secondary resonances.

The second critical feature of the theory is the relation between resonance frequency and polymer length. It is important to emphasize that this relation is not postulated, but rather is a consequence of the quantum dynamics of phonon super-radiance and hence of the quantum-limited amplification process which must occur if we are to understand quantum-limited detection. Again, a number of experiments support this prediction. In the alligator lizard,<sup>145</sup> the spiny granite lizard,<sup>146</sup> the chick,<sup>147</sup> and the chinchilla,<sup>148</sup> it has been found that receptor cells which are tuned to

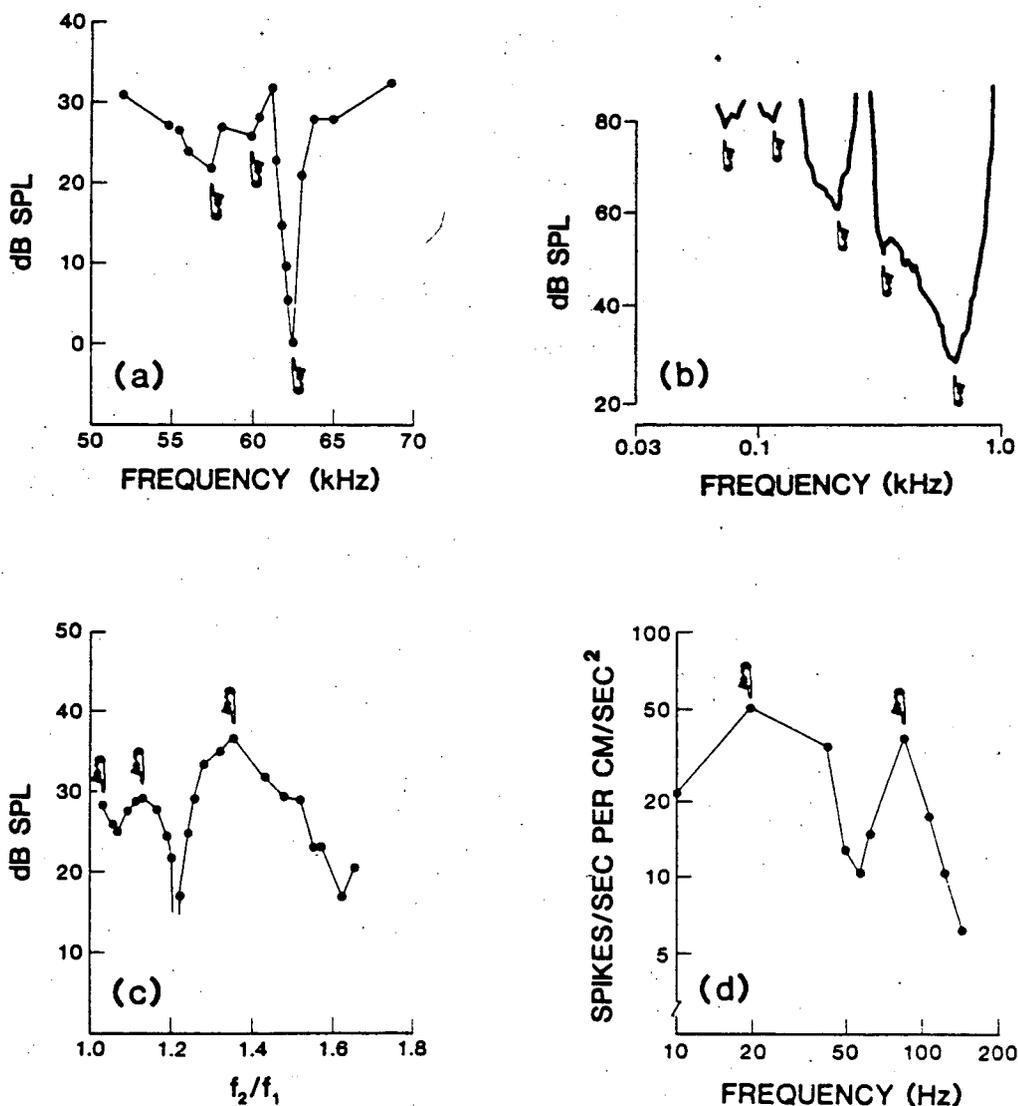


FIGURE III-25 Multiple resonances in the filters of the inner ear--experiment. (a) Extracellular receptor potential in the bat *Chilonycteris parnelli parnelli*.<sup>144</sup> Sound pressure level required to elicit a fixed voltage response to a pure tone, as a function of frequency. (b) Single neuron recordings in the alligator lizard *Gerrhonotus multicarinatus*.<sup>141</sup> Sound pressure level required to produce a fixed change in the rate of neural firing, as a function of frequency for a sinusoidal sound pressure at the eardrum. (c) Sound pressure in the ear canal of a cat.<sup>143</sup> Intensity of the combination tone  $2f_1 - f_2$  as a function of the frequency ratio  $f_1/f_2$  of the two primary tones at  $f_1$  and  $f_2$ ; in general different results are obtained for different primary intensities and primary frequencies, but the qualitative features, including multiple resonances, remain fixed. (d) Single neuron recordings from the frog sacculus.<sup>142</sup> Change in instantaneous neural firing rate per unit acceleration of the whole frog as a function of frequency for sinusoidal accelerations. In each case, "secondary resonances" are marked by arrows.

lower characteristic frequencies possess longer stereocilia. In the alligator lizard it is known that this receptor cell tuning does not reflect tuning of the basilar membrane, which moves with a frequency response essentially the same as that of the middle ear bones.<sup>149</sup> In the alligator lizard, and in the spiny granite lizard as well, the stereocilia stand freely in the fluid over a large portion of the organ, so that tuning is not related to the mechanics of auxiliary structures but to the stereocilia and receptor cells themselves.

In contrast to the interpretation presented here, the correlation between cilium length and resonance frequency has previously been taken as evidence for a simple passive mechanical resonance of the stereocilia, perhaps in concert with auxiliary structures in those organs where such are present.<sup>145</sup> We have seen, however, that such an interpretation is untenable--the small mass of the stereociliary bundle requires a very small stiffness to generate a resonance in the audible range and this small stiffness leads to enormous levels of Brownian motion. Furthermore, the Brownian motion of a passively resonant stereocilium would be concentrated in a narrow band of frequencies around resonance, and the reduction of this noise to tolerable levels would therefore require unreasonably long integration times in the receptor cell or central nervous system. Finally, the evidence that the stereocilium is hinged at its base<sup>136,150</sup> implies that, even if an effective mass is synthesized by some feedback process within the cell--perhaps coupled to electrical events at the cell membrane, as suggested<sup>151-153</sup> by several authors--the cilium stiffness will not depend significantly on polymer length, so that the relation between resonance frequency and cilium length could not be understood even in terms of a classical feedback model. In contrast, the quantum mechanical theory presented here provides a natural explanation of this correlation, albeit perhaps not an immediately obvious one.

A second problem in understanding the correlation between cilium length and characteristic frequency has been the small range of variation observed for the former,

particularly in mammals. The extreme example is that the stereocilia in the cochleae of bats are not terribly different from those in "normal" mammals,<sup>154</sup> although the resonance frequencies may differ by an order of magnitude; similar discrepancies exist regarding the variations within one animal.<sup>148</sup> Again these discrepancies argue against a simple interpretation in terms of passive mechanical resonances, which would vary as some small power of the cilium length.

In the theory presented here, the resonance frequency of the stereocilium is not completely determined by polymer length. Rather the polymer length defines a set of *possible* resonance frequencies, from which a dominant resonance is chosen by the strength of the chemical pump within the receptor cell. The same length stereocilium, simply by expending more energy, can therefore be tuned to much higher resonance frequencies, corresponding to higher index  $n$  in the resonance condition of Eq. (III.D.2.15).

The notion that, for example, receptor cells in the bat are distinguished from those in ordinary mammals by the value of  $n$  has two predictions. First, the large value of  $n$  implies that the secondary resonances will be closely spaced, since they are separated by  $\approx \Omega_r/n$ , with  $\Omega_r$  the dominant resonance. Second, the phenomenological model of feedback with a propagation delay leads to the prediction that the slopes of the resonance curves, expressed as dB/octave, will be proportional to  $n$ ,  $S \approx 20n$  dB/octave; this sharper response for larger  $n$  is analogous to the increased sharpness of interference patterns in the presence of multiple reflections, e.g. in a Fabry-Perot interferometer.<sup>155</sup> Thus there is an approximate relation between the spacing of the secondary resonances and the tuning slopes, which should be observed most dramatically in very high frequency cells with large values of  $n$ . Indeed, the frequency responses of bat auditory neurons are almost unmeasurably sharp, having been estimated to be  $\approx 3000$  dB/octave.<sup>156</sup> This suggests that secondary resonances will be very closely spaced in the bat, and this is confirmed at least qualitatively by the

experiments shown in Fig. III-25.

Several aspects of the quantum mechanical theory proposed here are thus in good agreement with experiment, and further studies of these effects should provide more detailed tests of the theory. In addition, it is possible to directly test the hypothesis that the mechanics of the stereocilium include feedback from some amplification mechanism, independent of the nature of the amplifier. Any mechanical system which is near to equilibrium will exhibit thermal noise spectra which are related to the response functions by the fluctuation-dissipation theorem.<sup>157</sup> In the presence of an amplifier--which drives the system away from equilibrium by virtue of its power supply--this theorem will not be obeyed. In effect the thermal noise which results from the equilibrium exchange of energy between the system and the heat bath is amplified, as shown in Fig. III-26, and this leads to the violation of the fluctuation-dissipation theorem.

If the displacement of the stereocilium can be measured both in response to controlled forces and thermal noise (spontaneous random displacements), then the comparison of these two measurements allows an unambiguous separation of active and passive components in the mechanical response. This method, although it separates the two components of the response, does not require any external perturbation of the system, for example metabolic poisons to remove the energy supply which drives the active response. Such direct measurements of stereocilium displacement and thermal noise should be accessible to the next generation of optical instrumentation for inner ear experiments, as are now being constructed in Berkeley and elsewhere.

The thermal noise of the stereocilium, and hence the contribution of active filtering to stereocilium mechanics, can be measured indirectly by its reflection in the firing pattern of single neurons or the voltage noise of individual receptor cells. Preliminary results of such analyses<sup>158</sup> are promising, but again systematic experiments will have to be undertaken to decide the issues. In particular it will be important to study the

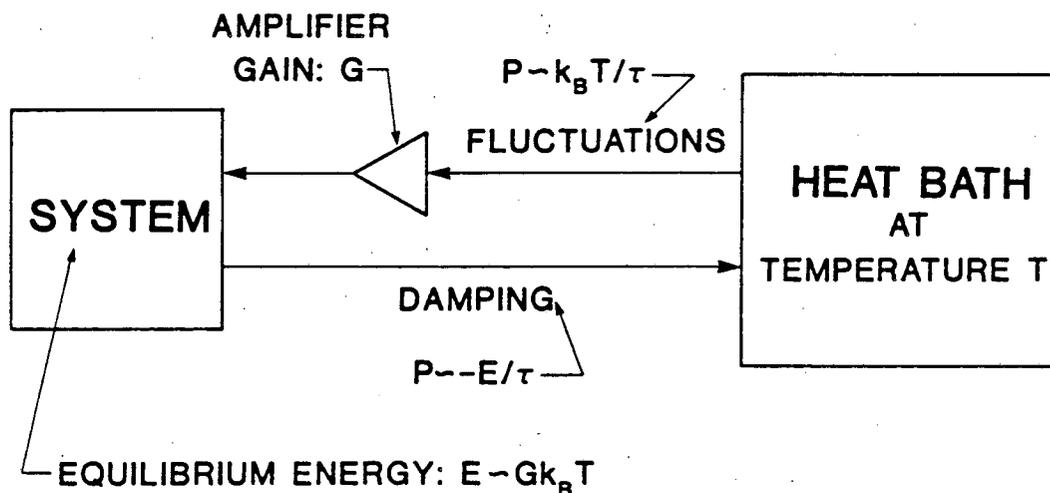
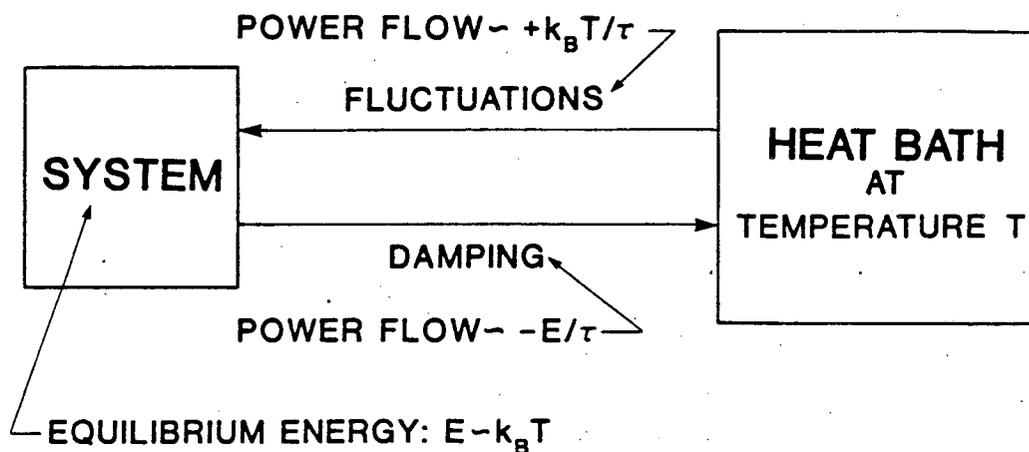


FIGURE III-26 Thermal noise and energy flow in active and passive systems, as described in the text.

variations with temperature and (if possible) fluid viscosity which these various measures of stereocilium noise exhibit, since it is only by such variations that the contribution of mechanical (as opposed to electrical, for example) noise sources can be identified unambiguously.

Finally, the contribution of active filtering to the mechanics of the inner ear may be detected by its behavior under pathological conditions. In any system with feedback through an amplifier, an increase in gain of the amplifier beyond some critical level can result in an instability of the system. Thus in a mechanical system such as the insect flight muscle and wing, activation of the muscle beyond a certain point results in spontaneous beating of the wing.<sup>159</sup> We expect that, under normal conditions, the amplifier within the stereocilium is biased away from such an instability, but since the gain of the amplifier is controlled by the internal chemistry of the receptor cell disorders of this chemistry can result in the transition to instability. Under these circumstances we should observe spontaneous "beating" of the stereocilium at a frequency close to its resonance when stable, and this beating will be reflected as sound *emitted* by the ear; the idea that sound emission by the ear could arise from instabilities of an active filter was first proposed in 1948 by Gold.<sup>152</sup>

Indeed a number of observations in several species, including humans, confirm the existence of such acoustic emissions.<sup>160-163</sup> Perhaps the most significant observation is that of Palmer and Wilson,<sup>163</sup> namely that such emissions occur from the inner ear of the frog at frequencies which correspond to the sensitivity of the basilar papilla, an organ where the stereocilia stand freely in the fluid. Thus, if acoustic emissions do reflect the instability of an active filter, then this filter must reside in the stereocilia themselves and not in auxiliary structures such as the basilar membrane; this is in accord with the conclusions drawn in Section II.C.3 from the analysis of thermal noise, and in contrast to models proposed by other workers.<sup>153,164</sup>

Although many features of the acoustic emissions, such as their phase-locking to external small amplitude stimuli<sup>165</sup> and their abrupt suppression by stronger stimuli,<sup>160,166</sup> are in accord with predictions from a model of an unstable active filter,<sup>167</sup> none of the existing experiments provide a decisive test of such a model. For example, the emission could reflect some extraneous (non-thermal) source of

noise which is filtered by the ear and thus appears as a narrow-band emission. Instability, which is indicative of an active element in the inner ear, can be distinguished statistically even if its average behavior can be simulated by a stable system. Thus, although the power spectrum of the acoustic emission may not allow us to distinguish an unstable filter from a stable filtered noise source, the probability distributions for the emission amplitude will be very different in the two cases.

In a system where zero amplitude is a stable point, the probability distribution for the amplitude must have a local maximum at this point, and will generally be roughly Gaussian. In contrast, a system for which zero amplitude is an unstable point will have a local minimum in the probability distribution at zero, and there will be two local maxima at  $\pm A$ , where  $A$  is the average amplitude of the emission. These differences, which are analogous to the change in probability distribution of an order parameter above and below the critical temperature<sup>157</sup> (stable and unstable, respectively) for a phase transition, are illustrated in Fig. III-27. Direct measurements of the probability distribution of acoustic emissions from either the frog or human ear are feasible, and these should provide clear evidence for or against the active filter hypothesis.

In summary, the explicitly quantum mechanical theory of the detector mechanism of the inner ear which I present here appears to be in accord with experiment wherever meaningful comparisons are possible. Most importantly, this mechanism is testable in detail by conventional anatomical, physiological, and mechanical experiments, in spite of the fact that its basic postulates are at the molecular scale and are expressed in a model Hamiltonian for the polymers which make up the stereocilium. Ideally, we imagine the following sequence of experiments:

- [1] Analysis of neural firing patterns and statistical behavior of acoustic emissions are used to find evidence for active components in the mechanical response of the stereocilium.

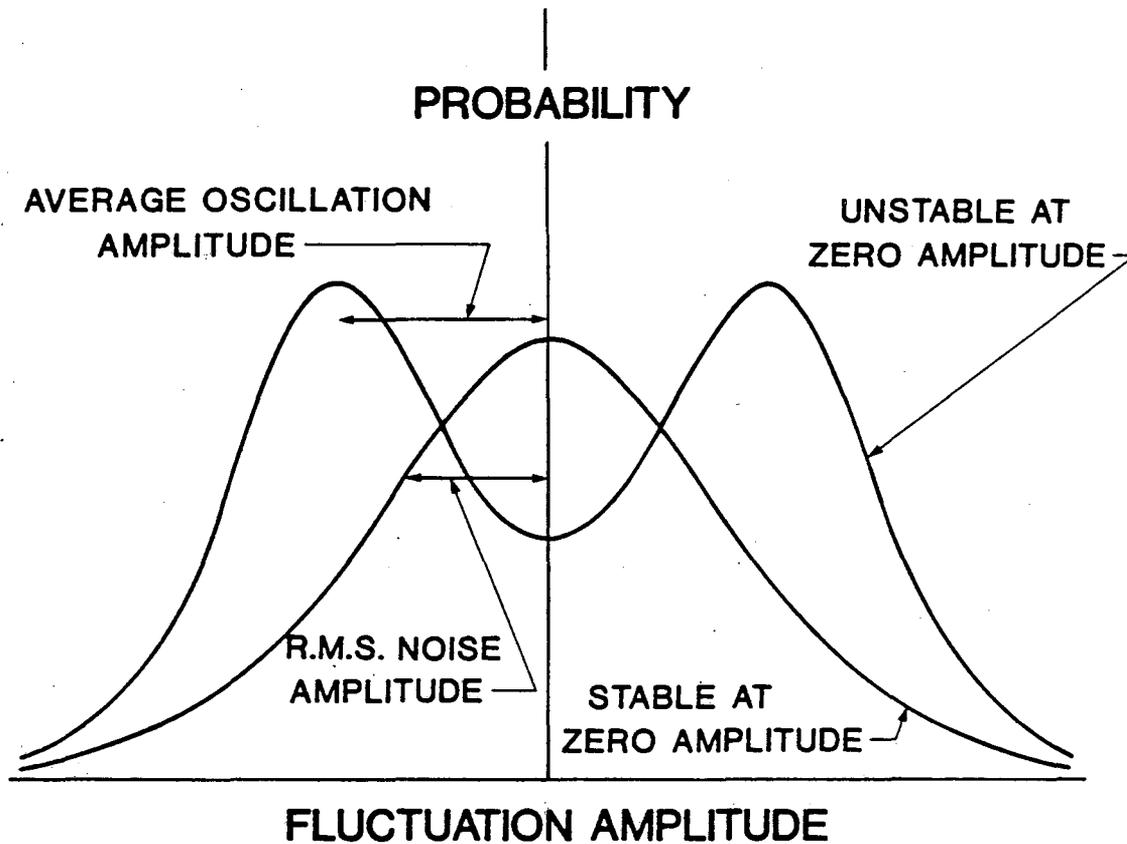


FIGURE III-27 Probability distributions for amplitude fluctuations in stable (e.g. a very narrow-band active filter) and unstable (e.g. an active filter with excessively high gain in its amplifier--an oscillator) systems.

- [2] The active and passive components in stereocilium mechanics are separated by testing for violations of the fluctuation-dissipation theorem in direct measurements of ciliary motion.
- [3] The "self-energy function" which characterizes the active component of stereocilium mechanics is examined for the resonant structure predicted from the properties of phonon super-radiance.
- [4] Variations in self-energy among receptor cells with different length stereocilia are compared with theory.<sup>168</sup>

The extant evidence strongly points to positive results in each case, although the most convincing possible studies have not yet been done. If all these tests can be passed,

there will be little doubt that the theory proposed here is correct at least in outline. For the present, the most significant conclusion is that an explicitly quantum mechanical theory of this macroscopic biological phenomenon of hearing--which is necessary if we are to make sense of the sub-angstrom motions which are detected at threshold--is subject to decisive experimental test with existing techniques.

## Chapter Four

### Concluding remarks: Where do we stand?

*...It is a trite saying that 'analogies cannot be pushed too far'...*

*W. Heisenberg, 1930<sup>1</sup>*

The realization that biological systems are subject to the same laws of physics and chemistry as inanimate objects marked a major step in the development of scientific thought. In the hundred years since this step was taken, analogies between biological and non-biological systems have been central to the understanding of molecular events in biology: small synthetic molecules are used to model the structures and spectroscopic features associated with active sites in biological macromolecules,<sup>2</sup> and the empirical rules of organic and inorganic chemistry are used to evaluate the contributions of various factors to the rates and specificities of biochemical reactions.<sup>3,4</sup> The central assumption in this approach is that the *physical* properties of a biological molecule are determined strictly by the *chemical* interactions of its constituent atoms and can thus be understood by analogy with the simpler molecules of the non-biological world. The work reported in this thesis is intended as a constructive attack on this point of view, in the hope of convincing the reader that, like all analogies, the attempt to understand biological molecules in terms of conventional chemical concepts

has its limits. In addition, I believe that my analysis of these issues leads to at least an outline for a new point of view on the physics of biological systems, particularly on the role of quantum mechanics in the dynamics of these systems.

One salient fact about life as we know it is that it exists in the condensed phase. Cells are surrounded by a quasi-solid membrane matrix in which various protein components are inserted, the interior of a cell consists of a concentrated, probably non-ideal solution, and even the individual molecules in this solution are so large as to resemble small samples of an amorphous solid, possessing  $\approx 10^4$  atoms. In the past several years a number of investigators<sup>5</sup> have therefore applied the methods of condensed matter theory to biological systems, and the results of Chapter III are a major extension of this approach. Perhaps the most surprising result is that relatively simple families of model Hamiltonians provide a quantitative description of bio-molecular dynamics in a wide variety of systems and parameter regimes, both on macroscopic and microscopic scales.

The model Hamiltonians which describe biological molecules involve a small number of electronic states strongly coupled to the molecular vibrations. The vibrational modes are, to a reasonable approximation, harmonic, and they are damped by their interactions with the surrounding solvent or membrane. These model Hamiltonians are very different from the classic electron-phonon interaction problem in metals, described by the Bardeen-Pines Hamiltonian.<sup>6</sup> The Bardeen-Pines Hamiltonian contains only terms corresponding to single phonon processes; that is, to lowest order in perturbation theory, a change in electronic state is accompanied by the emission or absorption of a single phonon. Furthermore, the coupling constant describing the probability of this process is in a certain sense small: Migdal<sup>7</sup> has shown that the exact vertex part for the electron-phonon interaction, which represents an effective coupling constant modified by corrections from higher orders in perturbation theory, differs from the "bare" coupling constant by terms of order  $(m/M)^{1/2}$ , where  $m$  and  $M$  are

the electron and ion masses, respectively. Thus the significance of multi-phonon processes is small by the same factor which determines the validity of the Born-Oppenheimer approximation. In contrast, we have seen that electronic transitions--chemical reactions--in biological molecules are essentially multi-phonon processes.

Several problems in condensed matter theory do, however, involve strong electron-phonon coupling. Among these are the motion of electrons in polar<sup>8</sup> and molecular<sup>9</sup> crystals, as well as the transfer of electrons and energy among impurity sites in metals and semiconductors.<sup>10</sup> In recent years, theoretical interest in systems with strong electron-phonon coupling has been stimulated by the observation that such systems can support collective excitations,<sup>11</sup> and more generally by the application of modern field-theoretical techniques to otherwise intractable problems.<sup>12,13</sup> In this spirit, Chapter III presented a functional integral method for multi-phonon electronic transitions and used this method to construct an effective diagrammatic perturbation theory which is at each order exact in the electron-phonon coupling, using the electronic matrix elements as the expansion parameters.

The theoretical techniques developed in Chapter III allowed the elucidation of several essential quantum effects which arise in biological molecules, as well as defining the conditions under which these effects will be observed. The two major effects are a resonant dependence of reaction rates on the vibrational spectrum of the molecule and the persistence of electronic coherence in reactions which compete with vibrational relaxation,<sup>14</sup> and experiments on ligand binding in heme proteins<sup>15,16</sup> and the photo-induced electron transfer reactions in photosynthesis<sup>17</sup> were shown to provide strong evidence for these effects. From the analysis of these specific cases we may draw some general conclusions:

- [1] The kinetic behavior of biological molecules depends not only on electronic factors (matrix elements and energy gaps) and the structures of reactant, product, and "transition" states, but also on the vibrational spectrum of the molecule.

Small changes in the spectrum can change the reaction rate by many orders of magnitude.

This is strictly a quantum mechanical effect; we have seen (cf. Section III.A.1) that a semi-classical treatment reproduces the essential results of transition state theory, so that there is no critical dependence on the vibrational spectrum. In the case of myoglobin the critical vibrational modes are of very low frequency ( $\approx 20 \text{ cm}^{-1}$ ), corresponding to "breathing" motions of a large portion of the macromolecule. This means that the essential features of the kinetic behavior cannot be reproduced by a small molecule model compound, simply because the spectrum of the small molecule does not extend to such low frequencies. Because the relevant vibrations involve large portions of the molecule, their spectrum can be effected by structural changes--such as the binding of other ligands--at points well removed (by  $\approx 1 \text{ nm}$  in the case of myoglobin) from the active site.

The ability of ligand binding at one point on an enzyme to affect catalytic activity at another point is one of the central aspects of enzymatic regulation,<sup>18</sup> and thermodynamic evidence suggests that such ligand binding reactions induce large changes in the dynamics of the protein,<sup>19</sup> as required for the "vibrational regulation" mechanism to operate. Further evidence for the contribution of vibrational factors to the control of reaction rates comes from the bacterial enzyme *Streptomyces gresius* protease A (SGPA). High resolution crystallographic studies<sup>20,21</sup> of this molecule demonstrate that two substrates, differing only by a hydroxyl group, result in structures of the enzyme-substrate complex which are identical to within  $\pm 0.005 \text{ nm}$ . Nonetheless, these two complexes differ in their vibrational dynamics, as measured (qualitatively) by the large differences in the Debye-Waller factors between the two complexes, and the catalytic rates for the two substrates differ by a factor of two.<sup>21</sup>

The case of SGPA is not unique; essentially all enzymes which have been studied at high resolution exhibit relatively small changes in structure as they proceed through

their catalytic cycle,<sup>22,23</sup> while where data are available it appears that the vibrational dynamics change significantly.<sup>24</sup> While these considerations are qualitative, the example of myoglobin demonstrates that if quantitative data are available on the dynamics of the molecule then the reaction rate can be interpreted in quantitative detail, perhaps even to the point of *a priori* calculation. The development of Raman<sup>25,26</sup> and infrared<sup>27</sup> techniques to study the vibrational modes of critical atoms in the active sites of enzymes thus holds great promise for providing the sort of data which we require to reach a quantitative understanding of the relations between molecular dynamics and biological function. The central point of the work presented here is that the theoretical framework within which this data is analyzed must be explicitly quantum mechanical if we are to reproduce the essential features of the biological system.

[2] The behavior of biological molecules is not always described by a collection of states and rate constants for transitions among these states. Under certain conditions quantum coherence is long lived and invalidates the semi-classical kinetic approach.

A general quantum statistical system is described by its density matrix; the diagonal components of this matrix measure the populations of the states while the off-diagonal components measure the coherence among pairs of states. When the coherence is destroyed rapidly on the time scale of interest only the diagonal components of the density matrix need be retained, and we recover a description in terms of populations and rate constants for transitions among states, but if the coherence is long lived no such reduced formulation is valid. The example of the primary electron transfer events in photosynthesis demonstrates both that this coherent regime is accessible to biological systems and that experimental consequences of quantum coherence are detectable. Further, theoretical arguments show that the conditions for coherent behavior are much less stringent than previously realized, so that many biological systems may operate in this regime.

One consequence of coherent dynamics in biological systems is thus that a description in terms of chemical rate constants is incomplete. It has previously been asserted<sup>28</sup> that, by including a sufficiently large number of states and kinetic parameters any biological system can be described with complete rigor and as thoroughly as the extant data allow. My results clearly demonstrate that this is not the case: kinetic descriptions require rapid randomization of quantum mechanical phases, and the dynamics of electron transfer in photosynthesis are simply not consistent with this random phase approximation.

The failure of the random phase approximation may admittedly not be important in very many biological systems; the existing evidence does not allow us to decide this issue. The theoretical situation is, however, quite clear: coherent dynamics are accessible to biological molecules, and this fact forces us to describe biological systems in terms of explicitly quantum mechanical models, with careful attention to the interplay between coherent and dissipative processes. The most important consequence of coherent dynamics in biological systems is that it allows quantum mechanical effects to be expressed not only in the kinetic behavior of individual molecules but also in the macroscopic behavior of molecular aggregates.

The evidence for macroscopic quantum effects in biology was discussed in Chapter II. Again we may generalize the results:

- [3] Sensory systems which make amplitude and phase measurements reach the limits to measurement imposed by the uncertainty principle. Quantum-limited measurement is accomplished in spite of a much larger amount of thermal noise.

This fact in and of itself demonstrates that the sensory systems operate in a regime where the uncertainty principle is not negligible; we are not permitted to let  $\hbar \rightarrow 0$ . Equivalently we may say that the sensory systems are *not* semi-classical systems, in spite of their macroscopic dimensions. It is this macroscopic quantum behavior which invalidates all previous models for the dynamics of these systems, and it is this

behavior which forces us to describe the sensory process in explicitly quantum mechanical terms.

Even if the sensory systems did not reach the quantum limits, some mechanisms would be required to reduce thermal noise to tolerable levels. In the case of the inner ear, and also in other cases,<sup>29</sup> the only viable mechanism appears to be some form of active filtering. The phenomenon of frequency selectivity, which has long been the focus of research on hearing<sup>30,31</sup> and more recently other senses as well,<sup>32</sup> is thus seen to be essential not only in *analyzing* the incoming signal, but simply in *detecting* it against the ever-present background of thermal noise.

The fact that thermal and quantum limits to measurement are reached in the sensory systems provides a tremendous constraint on theories of the sensory detection process. When one measures the frequency response of a receptor in the inner ear, for example, one can always account for the results by postulating some effective mechanical or electrical properties for the various structures of that organ. Even if such a model is confirmed by direct measurement of these properties the question of their molecular origin need not be addressed. On the other hand, the noise level in the receptor will depend critically on the molecular basis of the filtering and detection, as illustrated by the example of the stimulus gated channel (Section II.C.4). In a sense the magnitude of the noise provides an absolute scale against which the consistency of any theory may be judged; this is particularly true in the case of quantum noise.

- [4] The approach to quantum-limited measurement requires a linear amplifier which reaches the optimum noise performance allowed by the uncertainty principle. Such devices must have microscopic mechanisms in which quantum coherence lives for a time comparable to the measurement time, on the order of one millisecond.

It is this result which completes the evidence in favor of Schrödinger's view of life as a macroscopic quantum phenomenon. There is simply no way to understand the detection of sub-angstrom displacements in the inner ear without postulating a molecular mechanism in which quantum coherence is manifest on macroscopic scales of time and distance. It is also this result which brings together the analysis of microscopic and macroscopic quantum effects in biology.

It is quite remarkable that a consistent understanding of a decidedly macroscopic phenomenon such as the perception of sound must necessarily involve a discussion of molecular processes. Even more surprising is the fact that these molecular processes cannot be described by the conventional language of chemical kinetics which has been so widely applied in biology. Thus on the one hand these results force closer links between the macroscopic and molecular events in biology, while on the other hand they force us to abandon some long-held notions about how the molecular events are to be analyzed. The resolution of this conflict lies in the detailed interpretation of macroscopic events in terms of quantum mechanical models for molecular dynamics, and I have presented a first step toward this goal in Section III.D.3.

After all this, where do we stand? I believe that my results leave no doubt as to the importance of quantum mechanics in the dynamics of biological systems. In extreme cases, such as quantum-limited detection in the sensory systems or the non-exponential kinetics of the heme proteins, quantum effects dominate to such an extent that classical approximations are useless and simple quantum mechanical models are much better starting points. It is difficult at this writing to know where on the scale from quantum to classical we shall find the "average" biological system, but the tacit assumption of classical behavior is clearly no longer warranted.

In each of the cases examined in this work, a rigorous quantum mechanical analysis has led not only to a consistent qualitative understanding of the biological facts but also to a series of quantitative predictions which may be tested by future

experiments. These analyses often demonstrate that seemingly obvious experimental results are subject to widely varying interpretations, but at the same time they define the new experiments which must be done to distinguish among the different possibilities. I hope that these results will stimulate a new generation of experiments which specifically test the existence of quantum mechanical effects in biological systems, and some suggestions for such experiments have been presented in the preceding sections.

## Appendices

### A. The sensory "threshold"

Many of the arguments in Chapter II are based on the comparison of some experimental estimate of the threshold signal with theoretical estimates of noise levels. While it is obvious that higher noise levels should correspond to higher thresholds, their quantitative relation is not obvious. For the special case of detecting a continuous signal in a continuous background of Gaussian noise, however, the theory of signal detectability<sup>1</sup> provides a rigorous interpretation of the threshold and its relation to the noise level.

Application of signal detection theory to behavioral and psychophysical experiments results in a measure of the sensory threshold which is unbiased and free of arbitrary assumptions regarding the strategy which the subject uses in detecting the stimulus.<sup>2</sup> The subject in a psychophysical experiment is viewed as making observations on a set of variables (*e.g.* the function describing sound pressure vs. time at the eardrum, or some quantity calculated from this function) which are drawn from one of two probability distributions, depending on whether the signal is present or not. "Noise" is represented by the fact that these probability distributions have non-zero variances; if the variances were zero, corresponding to no noise at all, arbitrarily small signals could be detected: there is no classical "threshold" below which the signal can never be detected. Furthermore, even with the noise arbitrarily small signals can be detected if the observer is willing to make frequent errors (confusing signal and no-signal trials), while for fixed error rate trades can be made between the frequency of

“missing” signals and of falsely identifying no-signal trials as containing the signal (“false alarms”).

By controlling these aspects of the observer's detection strategy, the experimenter can define threshold as the signal level required to reach a criterion level of *reliability* in the discrimination of signal from no-signal (noise). The threshold defined in this manner is such that the signal is equal in magnitude to the standard deviation of the noise for the conventional criterion of 76% correct. Behavioral thresholds thus measure the noise level directly: a behavioral threshold sound pressure  $P$  implies that the noise level in the auditory receptor is equal to the signal level produced by a tone of amplitude  $P$ .

Physiological experiments at the level of the receptor cell support the signal-detectability interpretation of the sensory threshold. The intracellular voltage response of the receptor to pure tones is either proportional to or quadratically dependent upon the sound pressure at the eardrum,<sup>3,4</sup> so that arbitrarily small signals will produce arbitrarily small responses: again, there is no classical “threshold.”

In the case of the turtle,<sup>3</sup> where the intracellular voltage noise has been analyzed in detail, the threshold for reliable detection in one receptor cell is within a few dB of the behavioral threshold of the animal. Thus the threshold signal can be reliably detected by examining the response of only one cell.

Beyond the receptor cell, the sensory signal is carried by a pattern of discrete events (neural impulses). This discrete character implies that the conventional results of signal detection theory, which are restricted to continuous signals, cannot be applied to the neuron. Unfortunately, the existence of discrete neural events is suggestive of the classical “threshold” concept, according to which we would identify the threshold signal as that required to produce one extra impulse above the background of spontaneous activity. While this “extra spike” criterion provides a convenient experimental measure of neural sensitivity, it has nothing to do with the threshold for reliable

detection or the noise level in the receptor cell.

To give a rigorous definition of threshold at the sensory neuron, we must provide a statistical characterization of neural firing. The firing is in fact a stochastic process, being approximately Poisson<sup>5,6</sup>; acoustic stimuli modulate the instantaneous rate  $r(t)$  of this process. For small amplitude signals with frequencies  $\omega/2\pi$  below  $\approx 6$  kHz in cat, the firing rate can be written

$$r(t) = r_0[1 + S\cos(\omega t + \phi)], \quad (\text{A.1})$$

where  $r_0$  is the mean firing rate in the absence of a signal and the phase-locking coefficient  $S$  is proportional to sound pressure from the smallest signals studied up to  $\approx 30$  dB above the behavioral threshold.<sup>7,8</sup> As in the case of the receptor cell, we see that arbitrarily small signals produce arbitrarily small responses, so that in spite of the discrete nature of neural firing there is no classical "threshold."

The threshold for reliable discrimination of signal from noise in the neuron may be calculated as follows. We consider the observation of a single neuron over the time interval  $0 < t < T$ . The results of one such observation are completely described by the set of times  $\{t_i\}$  at which neural impulses occurred; we refer to each set  $\{t_i\}$  as a realization of the stochastic process whose statistics are determined by the instantaneous firing rate  $r(t)$ . Assuming that the Poisson approximation is valid, the probability of observing  $n$  events in the time interval  $0 < t < T$  is given by

$$P(n; T) = e^{-Q} \frac{Q^n}{n!}, \quad (\text{A.2})$$

where

$$Q = \int_0^T dt r(t). \quad (\text{A.3})$$

Given that  $n$  events occurred, the conditional probability of observing the particular

realization  $\{t_1, t_2, \dots, t_n\}$  is

$$P(t_1, t_2, \dots, t_n | n; T) = Q^{-n} \prod_{i=1}^n r(t_i). \quad (\text{A.4})$$

Having observed the realization  $\{t_i\}$ , our task is to decide whether it arose from the application of a stimulus or simply from the random spontaneous firing. The two possible conditions--signal and no signal (noise)--are described by the instantaneous firing rates  $r_+(t)$  and  $r_-(t)$ , respectively; each instantaneous rate function determines different probability distributions  $P_{\pm}(n; T)$  and  $P_{\pm}(t_1, t_2, \dots, t_n | n; T)$ . From these distributions we can form a "likelihood ratio"<sup>2</sup> which measures the relative probability that a given realization  $\{t_i\}$  arose from signal or noise:

$$\lambda(t_1, t_2, \dots, t_n) = \ln \left\{ \frac{P_+(n; T) P_+(t_1, t_2, \dots, t_n | n; T)}{P_-(n; T) P_-(t_1, t_2, \dots, t_n | n; T)} \right\} \quad (\text{A.5})$$

$$= Q_- - Q_+ + \sum_{i=1}^n \ln[r_+(t_i)/r_-(t_i)]. \quad (\text{A.6})$$

Note that  $\lambda$  ranges from  $-\infty$  to  $+\infty$ , and that  $\lambda > 0$  corresponds to signal being more likely than noise.

To determine the reliability of discriminating signal from noise, we require the probability distributions  $P_{\pm}(\lambda)$  for the likelihood ratio in both signal and noise conditions<sup>2</sup> To obtain these distributions, we must sum over all possible numbers of events  $n$ , while for each  $n$  we must integrate over all possible realizations  $\{t_1, t_2, \dots, t_n\}$ , but these integrations must be constrained to include only those realizations which correspond to a particular value of  $\lambda$ . Thus we obtain

$$P_{\pm}(\lambda) = \sum_{n=0}^{\infty} P_{\pm}(n; T) \int_0^T dt_1 \int_0^T dt_2 \dots \int_0^T dt_n P_{\pm}(t_1, t_2, \dots, t_n | n; T) \\ \times \delta[\lambda - \lambda(t_1, t_2, \dots, t_n)] \quad (\text{A.7})$$

$$\begin{aligned}
 &= \sum_{n=0}^{\infty} P_{\pm}(n; T) \int_0^T dt_1 \int_0^T dt_2 \cdots \int_0^T P_{\pm}(t_1, t_2, \dots, t_n | n; T) \\
 &\quad \times \delta(\lambda - Q_- + Q_+ - \sum_{i=1}^n \ln[r_+(t_i)/r_-(t_i)]). \tag{A.8}
 \end{aligned}$$

This expression can be simplified by using the Fourier representation of the delta function,

$$\delta(s) = \int \frac{dx}{2\pi} e^{ixs},$$

where  $x$  is a dummy variable. Substituting into Eq. (A.7), it may be shown that the multiple integrals over  $t_1, t_2, \dots, t_n$  reduce to  $n$  independent identical integrals, which allows the sum to be done exactly. The result is

$$P_{\pm}(\lambda) = e^{-Q_{\pm}} \int \frac{dx}{2\pi} e^{ix(\lambda - Q_- + Q_+)} e^{\zeta_{\pm}(x)}, \tag{A.9}$$

where

$$\zeta_{\pm}(x) = \int_0^T dt r_{\pm}(t) e^{-ix \ln[r_+(t)/r_-(t)]}. \tag{A.10}$$

Given experimental measures of the instantaneous firing rate for the two stimulus conditions these equations allow us to directly compute the probability distributions for the likelihood ratio, after which the standard results of signal detection theory<sup>2</sup> may be applied to calculate the reliability of the discrimination between the two conditions. Note that the discrete nature of the neural signal, which prevented the straightforward application of detection theory, has been eliminated in favor the continuous variable  $\lambda$ .

For the particular case of detecting a small amplitude pure tone, where  $r_+$  is given by Eq. (A.1), we find the simple results

$$P_+(\lambda) = (4\pi r_0 TS^2)^{-1/2} \exp[-1/2(2r_0 TS^2)^{-1}(\lambda - 2r_0 TS^2)^2], \text{ and} \quad (\text{A.11a})$$

$$P_-(\lambda) = (4\pi r_0 TS^2)^{-1/2} \exp[-1/2(2r_0 TS^2)^{-1}\lambda^2]. \quad (\text{A.11b})$$

Thus the likelihood ratio is a Gaussian random variable of variance  $\sigma^2 = 2r_0 TS^2$ , and its mean changes by  $2r_0 TS^2$  when we apply the signal. The ratio of the change in mean to the standard deviation of the "noise" measures the detectability  $d'$  of the signal (cf. Ref. 2), so that in this case  $d' = S(2r_0 T)^{1/2}$ . The behavior of the detectability is quite reasonable: it increases in direct proportion to the the synchronization coefficient  $S$  and in proportion to the square root of the mean number of impulses  $r_0 T$  in the observation interval. Finally, a signal-to-noise ratio of unity requires<sup>9</sup>  $S = (2r_0 T)^{-1/2}$ .

We apply these results to the cat cochlear nerve fiber tuned to 800 Hz which was studied by Johnson.<sup>8</sup> This cell had  $r_0 = 60 \text{ s}^{-1}$ , while the behavioral integration time<sup>10</sup> at this frequency in the cat is  $T \geq 140 \text{ ms}$ . The threshold for reliable detection is therefore  $S \approx 0.24$ , and Johnson's data demonstrate that this level of synchronization is achieved for sound pressures below  $-5 \text{ dB SPL}$  at the eardrum, which is essentially equal to the behavioral threshold of the animal as calculated<sup>11</sup> from the data of Miller *et al.*<sup>12</sup> and Wiener *et al.*<sup>13</sup> As in the case of single receptor cells in the turtle, we conclude that the threshold signal may be reliably detected by examining the responses of only one nerve fiber, as promised in the text.

### B. Further comments on photon counting

In the absence of light, the current generated by vertebrate photoreceptors exhibits two types of fluctuation.<sup>14</sup> One is a Gaussian noise with spectral density

$$S(\omega) = \frac{S(0)}{(1 + (\omega\tau)^2)^2}, \quad (\text{B.1})$$

while the other consists of spontaneous photon-like events with time course

$$J_0(t) = A (t/\tau)^3 e^{-t/\tau}, \quad (\text{B.2})$$

where for any particular cell the value of  $\tau$  determined from noise analysis is the same as that characterizing the single-photon response.

To calculate the detectability of single photons, we begin by determining the discriminability of discrete photon-like events from the continuous noise background. As described in Section II.A.3, this calculation reduces to evaluating

$$M = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_0^{\infty} dt_1 \int_0^{\infty} dt_2 \frac{e^{i\omega(t_1-t_2)}}{S(\omega)} J_0(t_1) J_0(t_2). \quad (\text{B.3})$$

Noting that

$$\int_0^{\infty} dt e^{i\omega t} J_0(t) = \frac{6A\tau}{(1-i\omega\tau)^4}, \quad (\text{B.4})$$

we see that

$$M = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{36A^2\tau^2}{S(0)} \frac{1}{(1+(\omega\tau)^2)^2} = \frac{9A^2\tau}{S(0)}. \quad (\text{B.5})$$

Typical parameter values are given in the papers of Baylor and co-workers,<sup>14,15</sup> from which we find  $M \approx 10$  promised in the text, so that the probability of reliably discriminating a photon-like event from continuous background noise is essentially one.

Given that photon-like events are reliably discriminable from the continuous noise, the detection problem reduces to deciding whether a given event was spontaneous or the result of photon absorption. In the absence of light the spontaneous events occur at some rate  $I_0$ . When we have a light of intensity such that the receptor absorbs  $I$  photons per second, the events still occur in a Poisson stream (assuming that the light source has Poisson statistics), but at a new rate  $I_0 + I$ . The problem of discriminating between these two Poisson sequences can be solved using the methods

of the preceding Appendix, the result being that a signal-to-noise ratio of unity is achieved when  $I \approx \sqrt{I_0 T}$ , where  $T$  is the integration time.

Given that the occurrence of spontaneous photon-like events is limiting at low intensities, it is of some interest to determine the origin of these events. One unavoidable source of noise is the ever-present blackbody radiation. Thus at frequency  $\Omega$ , a unit volume at temperature  $T$  contains

$$n(\Omega) = \frac{\Omega^2 d\Omega}{2\pi^2 c^3} \frac{1}{e^{\hbar\Omega/k_B T} - 1} \quad (\text{B.6})$$

photons in a unit bandwidth  $d\Omega$ , where  $c$  is the velocity of light. If the visual pigments of a receptor cell have an absorption cross-section  $\sigma(\Omega)$ , then the rate at which blackbody quanta will be detected is

$$R = c \int d\Omega \sigma(\Omega) n(\Omega). \quad (\text{B.7})$$

As a simple approximation, consider a Gaussian form for  $\sigma(\Omega)$ ,

$$\sigma(\Omega) = \sigma_0 e^{-(\Omega - \Omega_0)^2 / 2(\Delta\Omega)^2}. \quad (\text{B.8})$$

We know that  $\Omega_0 \gg k_B T / \hbar$  for visible light, so that the dominant contribution to the integral in Eq. (B.7) comes from frequencies where  $\sigma \approx \sigma_0 e^{-1/2(\Omega_0/\Delta\Omega)^2}$ . With this approximation, which underestimates  $R$ , we obtain

$$R \approx \frac{k_B T}{\hbar} e^{-1/2(\Omega_0/\Delta\Omega)^2} \frac{1}{6\pi^2} \sigma_0 (k_B T / \hbar c)^2. \quad (\text{B.9})$$

For rod cells in the toad retina,<sup>15,16</sup>  $\sigma_0 \approx 10 (\mu m)^2$  for the whole cell, and  $c/\Omega_0 \approx 500 \text{ nm}$ , while the observed value of  $R$  for these cells is<sup>14</sup>  $\approx 0.02 \text{ s}^{-1}$ . In order for the blackbody contribution of Eq. (B.9) to be smaller than the total, we require  $\Delta\Omega/\Omega_0 < 0.13$ , which amounts to an absorption bandwidth of  $\approx 70 \text{ nm}$ , very near to what is found experimentally.<sup>15</sup>

From these calculations we conclude that a single vertebrate photoreceptor cell is capable of counting individual photons, and that the reliability of this counting is limited only by thermal noise due to blackbody radiation.

### C. Mechanics of the cochlea

The treatment given in this Appendix follows portions of a review by Lewis, Leverenz, and the author<sup>17</sup>; other excellent reviews are available.<sup>18-20</sup> We idealize the cochlea as consisting of two straight chambers each of height (measured in the direction  $y$ )  $h$ . The two chambers are separated by the basilar membrane, which is assumed to be thin and thus serves only as a boundary condition at  $y = 0$ . The membrane is characterized by a mechanical impedance per unit area (pressure/point velocity)  $Z_{BM}(x;\omega)$  which varies smoothly with position  $x$  along the cochlea, and the details of fluid motion in the remaining  $z$  direction are neglected (two-dimensional approximation). Other than basilar membrane, all walls of the cochlea are rigid and hence stationary; all dimensions of the cochlea are small compared to the wavelengths of sound in the fluid at normal auditory frequencies, so that the fluid may be assumed incompressible. With these considerations in mind, the equations of motion for the system are as follows:

- [1] The Navier-Stokes equations for the fluid:

$$\rho \left( \frac{\partial}{\partial t} + \bar{v} \cdot \bar{\nabla} \right) \bar{v} = -\bar{\nabla} p + \eta \nabla^2 \bar{v}, \text{ and} \quad (\text{C.1})$$

$$\bar{\nabla} \cdot \bar{v} = 0, \quad (\text{C.2})$$

where  $\rho$  is the fluid density,  $\bar{v}$  is the fluid point velocity,  $p$  is the pressure, and  $\eta$  is the fluid viscosity.

- [2] The boundary conditions at the chamber walls:

$$\bar{v}(x, y = h) = \bar{v}(x, y = -h) = 0. \quad (\text{C.3})$$

[3] The boundary conditions at the basilar membrane:

$$\bar{v}(x, y=0) = \hat{y}V_{BM}, \text{ and} \quad (C.4)$$

$$p(x, y=0^+; \omega) - p(x, y=0^-) = Z_{BM}(x; \omega) V_{BM}(x; \omega), \quad (C.5)$$

where  $V_{BM}$  is the basilar membrane velocity.

The non-linear term in these equations,  $\bar{v} \cdot \bar{\nabla} \cdot \bar{v}$ , is small as long as the amplitude of fluid motion is small in comparison to its wavelength; this is always true in the cochlea at comfortable sound intensities. Once the equations are linearized in this way, we may define new variables  $p_{\pm}(x, y) = p(x, y) \pm p(x, -y)$  and similarly for  $\bar{v}_{\pm}$ . It is clear that only  $p_-$  and  $\bar{v}_-$  couple to the basilar membrane, and the remaining variables may be discarded.

The linear partial differential equations which relate  $p_-$ ,  $\bar{v}_-$ , and  $V_{BM}$  may be formally solved and converted to an integral equation which involves only the basilar membrane velocity; following de Boer<sup>21</sup> this may be termed the Siebert<sup>22</sup> equation:

$$V_{BM}(x; \omega) = \frac{1}{2\pi Z_{BM}(x; \omega)} \int dk \int dx' e^{ik(x-x')} Q(k; \omega) V_{BM}(x'; \omega), \quad (C.6)$$

where

$$Q(k; \omega) = -(2i\beta\rho\omega/k) \times \frac{e^{\beta h} [(\beta - k)e^{-kh} + (\beta + k)e^{+kh}] - e^{-\beta h} [(\beta + k)e^{-kh} + (\beta - k)e^{+kh}]}{(\beta + k)^2 [1 - \cosh((\beta + k)h)] - (\beta - k)^2 [1 - \cosh((\beta - k)h)]}, \quad (C.7)$$

$$\beta = [k^2 + i\omega/\nu]^{1/2}, \quad (C.8)$$

with  $\nu = \eta/\rho$  being the kinematic viscosity.

If the membrane impedance did not depend on position, these equations would be solved by running waves  $V_{BM} \approx e^{\pm ikx}$ , where the wavenumber  $k(\omega)$  is a solution of

$$Q(k; \omega) = Z_{BM}(\omega). \quad (C.9)$$

In the spirit of the WKB approximation, we may extend this solution to define a local wavenumber  $k(x; \omega)$  by

$$Q[k(x; \omega); \omega] = Z_{BM}(x; \omega). \quad (C.10)$$

At the entrance to the cochlea, the basilar membrane is very stiff and therefore present a large pure imaginary impedance. Large impedances correspond to small wavenumbers, as may be shown from Eq. (C.7); small wavenumbers correspond to allowing  $\nu \rightarrow 0$  in Eq. (C.8). Thus we begin by considering the long-wavelength, inviscid limit, where

$$k^2(x; \omega) = -\frac{2i\rho\omega}{hZ_{BM}(x; \omega)}. \quad (C.11)$$

For the case where  $Z$  is dominated by a stiffness  $K(x)$ , this becomes

$$k(x; \omega) = \omega[2\rho/hK(x)]^{1/2}. \quad (C.12)$$

Thus, energy flow into the cochlea begins with a non-dispersive wave whose wavelength shortens as it travels, since  $K(x)$  decreases. Careful analysis of the WKB prefactor<sup>23</sup> demonstrates that as the wavelength shortens the amplitude of basilar membrane motion increases, albeit slowly.

Eventually this simple pattern of propagation is terminated, as the membrane impedance crosses over to being resistance- and perhaps ultimately mass-dominated. The resulting attenuation of the traveling wave, which may be calculated from the imaginary part of  $k$ , can be quite large, so that energy of the wave does not propagate

very far past the cross-over point. The hydrodynamics of the cochlea thus serve to pull energy out of the fluid and onto the membrane, where it is dissipated in a small region; for detailed analyses of energy flow see the recent work of Lighthill<sup>20</sup> and de Boer.<sup>21</sup>

Some footnotes to this brief discussion:

[1] The function  $Q$  determines the relation between fluid velocity and membrane velocity. For the analysis of basilar membrane thermal noise *via* equipartition (Sect. II.C.3) we require this function in the inviscid limit, where it is independent of  $\omega$ .

[2] It is unlikely that the basilar membrane impedance is dominated by the mechanics of the basilar membrane itself. If, for example, the damping elements simply arose from internal viscosity of the membrane or interaction with the surrounding fluids, then all of the power in the incoming signal would be lost, and none of it delivered to the stereocilia. Since the threshold power is essentially equal to the thermal noise power in the experimentally observed detection bandwidths (cf. Sect. II.C.3), this is an impossible situation.

Consider instead that the local displacement of the membrane  $z_{BM}$  is coupled to the displacement of the stereocilia  $x_S$ . Then we will have

$$x_S(\omega) = \alpha_S(\omega) \left\{ F_S(\omega) + \Gamma(\omega) [X_{BM}(\omega) - x_S(\omega)] \right\}, \quad (C.13)$$

where  $\alpha_S$  is the stereocilium response function to forces  $F_S$  applied directly to the cilium and  $\Gamma$  is a coupling strength; similarly, by reciprocity,

$$X_{BM}(\omega) = [i\omega Z_{BM}(\omega)] \left\{ F_{BM}(\omega) + \Gamma(\omega) [x_S(\omega) - X_{BM}(\omega)] \right\}. \quad (C.14)$$

Solving these equations it may be shown that, in the absence of forces directly applied to the cilium, the membrane responds to applied forces with an effective impedance

$$Z_{BM}^{eff}(\omega) = Z_{BM}(\omega) + (i\omega)^{-1} \frac{\Gamma(\omega)}{1 + \Gamma(\omega)\alpha_S(\omega)}. \quad (C.15)$$

In the limit of strong coupling ( $\Gamma \rightarrow \infty$ ), we see that the basilar membrane directly reflects the impedance of the cilia, so that if the latter exhibit an active resonance, so will the *apparent* impedance of the membrane, as promised in the text.

#### D. Radiationless transitions *via* the Bloch equations

In the case of a two-state system, the model Hamiltonian can be written either in terms of fermion creation and annihilation operators or in terms of Pauli spin matrices, subject to the following identifications:

$$\sigma_z = c_1^\dagger c_1 - c_2^\dagger c_2 \quad (D.1a)$$

$$\sigma_+ = c_1^\dagger c_2 \quad \sigma_- = c_2^\dagger c_1. \quad (D.1b)$$

The Hamiltonian of Eq. (III.A.3.40) therefore becomes

$$\mathbf{H} = \sigma_z [\epsilon/2 + D] + VF\sigma_+ + VF^\dagger\sigma_- + \mathbf{H}_{ph} - \langle \sigma_z \rangle D, \quad (D.2)$$

where  $\epsilon = \epsilon_1 - \epsilon_2$  and similarly for  $D$ . Comparing this equation with Eq. (III.A.3.45), the stochastic coefficients of spin relaxation theory can be identified. From these identifications we obtain,<sup>24</sup> with  $\epsilon' = \epsilon + 2\langle D \rangle$ , the two time constants which characterize the time course of the electronic dynamics:

$$T_1^{-1} = \frac{2V^2}{\hbar^2} \int_{-\infty}^{\infty} d\tau e^{-\frac{\epsilon'}{\hbar}\tau} \langle F(\tau)F^\dagger(0) + F^\dagger(\tau)F(0) \rangle \quad (D.3)$$

$$T_2^{-1} = \frac{2}{\hbar^2} \int_{-\infty}^{\infty} d\tau [\langle D(\tau)D(0) \rangle - \langle D \rangle^2] + \frac{1}{2}T_1^{-1}. \quad (D.4)$$

To evaluate these formulae we require a method of computing correlation functions, such as  $V^2 \langle F^\dagger(\tau)F(0) \rangle$ . By analogy with the time-ordered generating function of Eq. (III.A.3.39), we have

$$\begin{aligned} & \langle :F_1(a_1^\dagger, a_1) :: F_2(a_2^\dagger, a_2) : \rangle = \\ & \int d^4\alpha d^4\xi e^{-i\bar{\xi}\cdot\bar{\alpha}} \langle e^{-i\xi_1 a_1^\dagger} e^{-i\xi_1 a_1} e^{-i\xi_2 a_2^\dagger} e^{-i\xi_2 a_2} \rangle \\ & \times :F_1(\alpha_1', \alpha_1) :: F_2(\alpha_2', \alpha_2) :, \end{aligned} \quad (D.5)$$

where  $a_1 = a(t_1)$ , etc, and

$$\begin{aligned} & \langle e^{-i\xi_1 a_1^\dagger} e^{-i\xi_1 a_1} e^{-i\xi_2 a_2^\dagger} e^{-i\xi_2 a_2} \rangle = \\ & \exp\{-\bar{\nu}(\xi_1 \xi_1' + \xi_2 \xi_2') - (\bar{\nu}+1)\xi_1 \xi_2' e^{i\omega'\tau - \gamma|\tau|} - \bar{\nu}\xi_1' \xi_2 e^{-i\omega'\tau - \gamma|\tau|}\}. \end{aligned} \quad (D.6)$$

For the harmonic approximation, this method is sufficient to cast the calculation of relevant correlation functions into two c-number Gaussian integrals over  $N$  four-dimensional variables, where  $N$  is the number of vibrational modes (recall that  $F$  is a Gaussian function of  $a$  and  $a^\dagger$ ). The Gaussian character of the generating functions stems from the harmonic approximation. If anharmonic effects are to be incorporated, higher-order Green's functions contribute higher cumulants to the generating functions. It should be stressed, however, that the *only* modification needed in the above method is in the generating function. In particular, the calculation of  $F$  need not be changed, although it may be convenient to do so.

When a system contains a only single vibrational mode and the frequency shift,  $\rho/\omega'$ , is small, then in the harmonic approximation we find:

$$T_1^{-1} = \frac{2V^2}{\hbar^2} \int_{-\infty}^{\infty} d\tau \exp\left\{-i(\epsilon/\hbar + (\bar{\nu}+1/2)\rho)\tau + \eta^2(x+y-2\bar{\nu}-1)\right\} \quad (D.7)$$

$$T_2^{-1} = \frac{2\rho^2}{\gamma} \bar{\nu}(\bar{\nu}+1) + 1/2 T_1^{-1}; \quad (D.8)$$

$$\beta = 1/2 \ln\left\{\frac{\omega-\rho}{\omega+\rho}\right\} \quad (D.9)$$

$$\eta = \Delta\{m\omega'/2\hbar\}^{1/2} \left\{ \langle \sigma_z \rangle \left( 1 - e^\beta \right) - \left( 1 + e^\beta \right) \right\} \quad (\text{D.10})$$

$$x = (\bar{\nu}+1) \exp\left\{ i\omega'\tau - \gamma|\tau| \right\} \quad (\text{D.11})$$

$$y = \bar{\nu} \exp\left\{ -i\omega'\tau - \gamma|\tau| \right\}. \quad (\text{D.12})$$

Before proceeding with the calculations, some interesting qualitative points may be noted. Loss of vibrational coherence ( $\gamma$ ), and electronic coherence ( $T_2$ ) occur on different time scales, and there is no general rule governing their relation. In the absence of vibrational-frequency changes between the two electronic states,  $T_2 = 2T_1$ , so that electronic coherence persists for the full time course of the radiationless transition. Frequency changes can reduce  $T_2$  considerably, but we always have  $T_2 = 2T_1$  in the low temperature limit, as may be seen from Eq. (D.8) This result, in conjunction with the condition for validity of perturbation theory given below, shows that  $T_2V/\hbar \gg 1$ . Thus, at low temperature or when frequency changes are very small, electronic coherence persists for many cycles of the "exchange time,"  $\hbar/V$ ; this result is independent of other features of the model, and is in agreement with experiment.<sup>25</sup>

Returning to the details, the integral in Eq. (D.7) is calculated by power series expanding in  $x$  and  $y$ , and directly integrating the resulting exponentials in  $\tau$  to obtain a sum of Lorentzians. If the effective energy gap  $\epsilon' = \epsilon + (\bar{\nu} + 1/2)\rho\hbar$  is an integer multiple of  $\hbar\omega'$ , we can use a Bessel function identity to sum the dominant terms:

$$T_1^{-1} = \frac{4V^2 e^{-S(2\bar{\nu}+1)}}{\hbar^2 \gamma} \left( \frac{\bar{\nu}+1}{\bar{\nu}} \right)^{\frac{\epsilon'}{2\hbar\omega'}} \int_0^1 dz z^{-1} I_{\frac{\epsilon'}{\hbar\omega'}} \left( 2Sz\sqrt{\bar{\nu}(\bar{\nu}+1)} \right) \quad (\text{D.13})$$

where  $S=\eta^2$ ,  $I$  is a modified Bessel function, and  $z$  is a dummy variable.<sup>26</sup>

To obtain the rate constant when the vibrational line is inhomogeneously broadened on a time scale faster than  $T_1$ , Eq. (D.7) must be convolved with the vibrational lineshape; if this is Gaussian, the following result obtains, neglecting terms in  $\rho/\omega'$ :

$$T_1^{-1} = \frac{4V^2 e^{-S(2\bar{\nu}+1)}}{\hbar^2 \gamma'} \frac{\hbar \omega'}{\epsilon'} \left( \frac{\pi}{2} \right)^{1/2} \left( \frac{\bar{\nu}+1}{\bar{\nu}} \right)^{2\bar{\nu}} I_{\frac{\epsilon'}{\hbar \omega'}} \left( 2S \sqrt{\bar{\nu}(\bar{\nu}+1)} \right), \quad (\text{D.14})$$

where  $\gamma' \gg \gamma$  is the inhomogeneous linewidth.

These results cannot be extended analytically to non-integral values of  $\epsilon'/\hbar\omega'$ , because  $\int_{-\infty}^{\infty} d\epsilon' T_1^{-1}(\epsilon')$  calculated from Eq. (D.13) would differ from the same quantity correctly calculated from Eq. (D.7). This limitation can be overcome by the following stratagem. Viewed as a function of  $\epsilon'$ ,  $T_1^{-1}$  is a sum of Lorentzians, whose peak positions are integral multiples of  $\hbar\omega'$ , and whose widths depend on  $\gamma$ . We approximate the sum of all Lorentzians with fixed peak position as a Gaussian whose variance will be determined such that  $\int T_1^{-1} d\epsilon' = 4\pi V^2/\hbar$ , as can be directly calculated from Eq. (D.7). For example, in the inhomogeneous case,

$$T_1^{-1} = \sum_{n=-\infty}^{\infty} \frac{4V^2 e^{-S(2\bar{\nu}+1)}}{\gamma' \hbar^2} \left( \frac{\bar{\nu}+1}{\bar{\nu}} \right)^{\frac{n}{2}} \frac{\hbar \omega'}{\epsilon'} \left( \frac{\pi}{2} \right)^{1/2} I_n \left( 2S \sqrt{\bar{\nu}(\bar{\nu}+1)} \right) \\ \times \exp \left\{ - \frac{(\epsilon'/\hbar - n\omega')^2}{2\sigma^2} \right\}, \quad (\text{D.15})$$

where  $\sigma$  is the standard deviation of the Gaussian. We find,

$$\sigma = 2\gamma'/S. \quad (\text{D.16})$$

These results may be generalized to the consideration of one low frequency mode and several high frequency modes. Consider a set of high-frequency modes with

frequencies  $\Omega_\mu$  and coupling constants  $S_\mu$ ; all these modes have  $\bar{\nu}=0$  since  $\Omega_\mu \gg k_B T/\hbar$ . If in addition there is one low-frequency mode with  $\omega$  and  $S$ , it may be shown that

$$T_1^{-1} = (2V/\hbar)^2 \sum_{n_\mu} \left[ \prod_\mu \frac{S_\mu^{n_\mu}}{n_\mu!} \right] \sum_{m,n} \frac{(S\bar{\nu})^n (S(\bar{\nu}+1))^m}{m!n!} \times \frac{2[\gamma(n+m) + \sum_\mu \gamma_\mu n_\mu]}{(\epsilon - \sum_\mu n_\mu \omega_\mu - n\omega)^2 + (n\gamma + \sum_\mu n_\mu \gamma_\mu)^2}, \quad (\text{D.17})$$

where the  $\gamma_\mu$  are damping constants for the high-frequency modes. This complicated expression may be simplified by considering a term with fixed  $n_\mu$ . Then for appropriate definitions of  $\Gamma$ ,  $\bar{\epsilon}$ , and  $V'$ , we have

$$T_1^{-1}(n_\mu) = (2V'/\hbar)^2 \sum_{m,n} \frac{(S\bar{\nu})^n (S(\bar{\nu}+1))^m}{m!n!} \frac{2(\gamma n + \Gamma)}{(\bar{\epsilon} - n\omega)^2 + (\gamma n + \Gamma)^2}, \quad (\text{D.18})$$

which is just the result for transitions with coupling to a single phonon mode, but with new energy gap and matrix element, as well as a finite natural lifetime  $\Gamma^{-1}$  for the electronic state; if we take the term with  $n_\mu=0$  then only the matrix element is modified.

To a reasonable approximation, we imagine that one of the terms in the sum over  $n_\mu$  will dominate. Evidently this term has the same form as that for a single-mode model, but with renormalized parameters, as promised in the text.

The results simplify once again in the limit that there are a large number of independent vibrational modes, so that integrand in Eq. (D.3) factors into a product of correlation functions, one for each mode:

$$T_1^{-1} = 2 \left( \frac{V}{\hbar} \right)^2 \int_{-\infty}^{\infty} d\tau \exp \left\{ -i \frac{\epsilon'}{\hbar} \tau \right\} \prod_j C_j(\tau) . \quad (\text{D.19})$$

If the modes have incommensurate frequencies,  $T_1$  can be evaluated by the central limit theorem.<sup>27</sup> The result is Gaussian, with a mean and variance given by the sum of means and sum of variances, respectively, of the Fourier transforms of the individual correlation functions. Thus,

$$T_1^{-1} = \frac{4\pi V^2}{\hbar^2 \sqrt{2\pi s}} \exp \left\{ -\frac{(\epsilon'/\hbar - M)^2}{2s} \right\} \quad (\text{D.20})$$

$$M = -i \sum_j \left. \frac{\partial \ln C_j(\tau)}{\partial \tau} \right|_{\tau=0} \quad \text{and} \quad s = -\sum_j \left. \frac{\partial^2 \ln C_j(\tau)}{\partial \tau^2} \right|_{\tau=0} . \quad (\text{D.21})$$

Within the Condon and harmonic approximations, the central limit theorem can still be applied even if the modes are not independent. From above, the correlation function can be written as an integral over  $N$  dummy variables, with  $N$  proportional to the number of modes; in the harmonic and Condon approximations this integral is Gaussian. It is always possible to bring an  $N$ -dimensional Gaussian integral into a product of  $N$  Gaussian integrals by a change of variables, and in these new variables the correlation-function factors as in Eq. (D.19). Thus the above method still applies. The Gaussian dependence of the reaction rate on the energy gap is called the energy gap law.<sup>28,29</sup> A number of workers<sup>28,30-32</sup> have found that, in some approximation, the energy gap law is valid, and this result is the basis for the interpretation of a large body of data on radiationless transitions.<sup>33</sup>

In the case of a single vibrational mode, however, Eq. (D.15) demonstrates that the energy-gap dependence of the transition rate has considerable structure, and is not well approximated by a Gaussian. This structure arises from resonances between the electronic energy gap and integer multiples of the vibrational frequency, and is not an interference effect as may occur with several modes.<sup>34</sup> Abram and Silbey<sup>35</sup> have noted

the existence of these resonances, but a proper treatment must involve explicit consideration of vibrational relaxation, as given here.

All of the considerations presented here are based on perturbation theory in the electronic matrix element,  $V$ . The condition that a perturbation series converge is that the ratio of successive terms be much smaller than unity. For the "slow-regime" theory of this Appendix, a rough argument is as follows. The first term in the reaction rate expansion is the product of  $(V/\hbar)^2$  and a single integral over a two-time correlation function. This results in  $(V/\hbar)^2$  times a characteristic correlation time,  $\tau_c$ , and in the limit of first-order perturbation theory, represents the reaction rate,  $T_1^{-1}$ . The second term is the product of  $(V/\hbar)^4$  and a triple integral over a four-time correlation function, resulting approximately in  $(V/\hbar)^4 \tau_c^3$ . Thus the expansion parameter is  $(V/\hbar)\tau_c$ , and the validity condition is  $1 \gg (V/\hbar)\tau_c$ , or  $1 \ll T_1 V/\hbar$ . Given  $T_1$ , the application of perturbation theory therefore requires  $V$  to be *greater* than a minimum value  $\hbar/T_1$ .

With these results in mind, we proceed to some considerations which are specific to the case of myoglobin treated in Sect. III.B. We are interested in a single mode model, for which Eq. (D.7) may be explicitly evaluated as

$$T_1^{-1} = k(\omega) = \frac{4V^2}{\hbar^2} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(S\bar{\nu})^m}{m!} \frac{(S(\bar{\nu}+1))^n}{n!} \\ \times \frac{\gamma(n+m)}{(\epsilon - (n-m)\omega)^2 + \gamma^2(n+m)^2}. \quad (\text{D.22})$$

The resonances as a function of  $\omega$  correspond to  $\omega_N = \epsilon/N$ , where  $N$  is an integer; the spacing between the  $N^{\text{th}}$  resonance and its neighbors is given approximately by  $\omega_N/N$ , so that for large  $\epsilon/\omega$  the resonances are quite closely spaced. In the case of myoglobin and carbon monoxide, the kinetic data<sup>36</sup> indicate  $\epsilon \approx 0.3 eV$ , while the arguments of the text suggest  $\omega \approx 20 \text{ cm}^{-1}$ . This implies that  $\epsilon/\omega \approx 120$ , so that

resonances are spaced by  $\approx 0.15 \text{ cm}^{-1}$ , which is very small indeed.

The close spacing of the resonances has two implications. First, it may be seen from Eq. (D.22) that the range of rate constants between a resonance and "half-way off" the resonance is  $\approx (\Delta\omega/\gamma)^2$ , where  $\Delta\omega$  is the frequency shift in moving half-way to the next resonance, or  $\approx 0.075 \text{ cm}^{-1}$  in the present case. To achieve the three order-of-magnitude range reflected in the extreme non-exponential time course of the ligand binding reaction therefore requires  $\gamma \approx 0.005 \text{ cm}^{-1}$ , which is exactly what we found in examining Brillouin scattering and Mössbauer data, as described in the text.

Second, the close spacing of the resonances implies that even a small amount of inhomogeneous broadening is sufficient to distribute molecules over several resonances, invalidating the single resonance approximation of Fig. III-13. We may make an alternative approximation, namely that the few resonances under the inhomogeneous lineshape are roughly identical and that all values of frequency from one resonance to the next are equally likely. This approach makes some small errors at the beginning and end of the time course, since extreme values of the rate constant are contributed by small variations from resonance to resonance, but should be sufficient for the rough agreement sought here.

Thus if we measure frequency in units of the resonance spacing, then for Lorentzian resonances of width  $\Delta$  and peak  $k_0$ , the time course of the reaction is

$$F(t) = \int_0^1 dx \exp \left\{ -\frac{k_0 t}{1 + (x/\Delta)^2} - \frac{k_0 t}{1 + (x/\Delta - 1/\Delta)^2} \right\}, \quad (\text{C.23})$$

where the second term indicates the effect of a the neighboring resonance. The effective Lorentzian width can be obtained from Eq. (D.22) by finding the dominant term of the series, and  $k_0$  is given by Eq. (D.13). It is this approximation which was used to generate Fig. III-15.

### E. Radiative interactions in strongly coupled systems

Almost all our information of the dynamics of biological molecules is derived from the interaction of these molecules with electromagnetic radiation. In some systems, such as vision and photosynthesis, these radiative interactions have functional significance. As in the case of radiationless transitions between the electronic states of biological molecules, we expect that the radiative interactions will be accompanied by the absorption or emission of several vibrational quanta--the interactions are "multi-phonon."

The theory of multi-phonon radiative processes is somewhat more subtle than the theory of the corresponding radiationless processes. Some of these subtleties do not seem to have been noticed in the literature, and this provided the motivation to develop the techniques presented here. As an example, consider the absorption of light by a two-level system strongly coupled to a single phonon mode; the absorption band consists of a sequence of "vibronic lines" separated by the phonon frequency. Each of these lines arises from specific vibronic state to vibronic state transitions, and each final state has a different radiative lifetime, so that the widths of the various absorption lines may be quite different. In practice, radiative damping has been included in existing theories only phenomenologically, as an apparent lifetime for the electronic state independent of the vibrational modes.<sup>37</sup>

Problems such as that of radiation damping can be avoided in a full Green's function approach such as that outlined here. I shall give a bare-bones version of a theory which is very rich and complex, and admittedly not completely worked out. The limited goal of this Appendix is to derive qualitative results for the absorption band of a coupled system which may be applied to the vibrational spectra discussed in Sect. III.B.2.

I begin with the simplest case, that in which the only matrix elements among electronic states are radiative in origin. If this is true, then the interaction

Hamiltonian may be written

$$\mathbf{H}_{int} = \sum_{ij} \sum_{\mathbf{k}\lambda} [g_{\mathbf{k}\lambda}^\dagger(ij) a_{\mathbf{k}\lambda} + g_{\mathbf{k}\lambda}(ij) a_{\mathbf{k}\lambda}^\dagger] c_i^\dagger c_j F_{ij}, \quad (\text{E.1})$$

where the  $F_{ij}$  arise from the canonical transformation of Eq. (III.A.3.3), with the proviso that  $F_{ii} = 1$ , the  $a_{\mathbf{k}\lambda}^\dagger$  create photons of wave-vector  $k$  and polarization  $\lambda$ , and the  $g_{\mathbf{k}\lambda}(ij)$  are a set of coupling constants. We are interested in the photons Green's functions, since these can ultimately be related to the scattering and absorption cross-sections. These functions are generated by the (un-normalized) generating functional

$$\Lambda'[J] = \langle \mathbf{T}[\exp\{-i \int dt \mathbf{H}_{int}(t) - iJ(t) \cdot a(t)\}] \rangle, \quad (\text{E.2})$$

where as before  $J(t) \cdot a(t) = \sum_{\mathbf{k}\lambda} \int dt [J_{\mathbf{k}\lambda}^\dagger(t) a_{\mathbf{k}\lambda}(t) + H.C.]$ .

Following the manipulations of Sect. III.A.3, we may write this generating functional in a functional integral representation:

$$\begin{aligned} \Lambda'[J] = & \int [d\alpha] \int [d\xi] \int [d\mu] \int [d\nu] e^{i\alpha \cdot \xi} e^{-iJ \cdot \alpha} e^{i\mu \cdot \nu} \\ & \times \langle \mathbf{T}[e^{-i\xi \cdot a}] \rangle \langle \mathbf{T}[e^{-i\nu \cdot c}] \rangle \langle \mathbf{T}[\exp\{-i \sum_{\mathbf{k}\lambda} \int dt [g_{\mathbf{k}\lambda}^\dagger(ij) a_{\mathbf{k}\lambda}(t) \\ & + H.C.] F_{ij}(t) \mu_i^\dagger(t) \mu_j(t)\}] \rangle. \end{aligned} \quad (\text{E.3})$$

We recall that  $\langle \mathbf{T}[e^{-i\nu \cdot c}] \rangle$  generates the non-interacting electron Green's functions, and is therefore Gaussian, by Wick's theorem. This implies that all the integrals over  $\mu$  and  $\nu$  are Gaussian and may be done exactly. The result, up to an irrelevant normalization constant, is

$$\Lambda'[J] = \int [d\alpha] \int [d\xi] e^{i\alpha \cdot \xi} e^{-iJ \cdot \alpha} \langle \mathbf{T}[e^{-i\xi \cdot a}] \rangle \langle \mathbf{T}[e^{Tr \ln \Xi}] \rangle, \quad (\text{E.4})$$

where the operator  $\Xi$  is defined by

$$\begin{aligned} \Xi_{ij}(t,t') &= \delta_{ij}\delta(t-t') \\ &+ \sum_{\mathbf{k}\lambda} \sum_l G_{il}^{(0)}(t-t') [g_{\mathbf{k}\lambda}^\dagger(lj)\alpha_{\mathbf{k}\lambda}(t') + H.C.] F_{lj}(t'). \end{aligned} \quad (\text{E.5})$$

Schematically we can expand  $Tr \ln \Xi$  in powers of the coupling constants,

$$Tr \ln \Xi = I + II + III + IV + \dots, \quad (\text{E.6})$$

where  $I \approx g$ ,  $II \approx g^2$ , etc.. To evaluate the time ordered expectation value, we may use the cumulant expansion,<sup>38</sup> which gives (again schematically)

$$\begin{aligned} \langle T[e^{Tr \ln \Xi}] \rangle &= \exp\{ \langle I \rangle \\ &+ [ \langle II \rangle + \frac{1}{2} \langle T[I \cdot I] \rangle - \frac{1}{2} \langle I \rangle \langle I \rangle ] + \dots \} \end{aligned} \quad (\text{E.7})$$

The first term is linear in both  $g$  and  $\alpha$ . When substituted into Eq. (E.4) this yields essentially a constant offset for the variable  $\alpha$ ; physically this corresponds to the electric field acquiring a constant mean value in response to the mean polarization the electronic states, which is an uninteresting effect from the point of view of spectroscopy.

The second term in Eq. (E.4) is quadratic in both  $g$  and  $\alpha$ , and therefore changes the photon Green's function without introducing non-Gaussian terms in the generating functional; this change endows each photon with a finite lifetime, corresponding to decay of the photon *via* absorption. Higher order terms related to *III* and *IV* (not shown) correspond to effective photon-photon interactions (non-Gaussian generating functional), and are responsible for the conversion of one photon into another (*e.g.* in Raman scattering).

The above considerations demonstrate that, so long as we are interested in lowest-order results for the photon self-energy, all that need be done is to evaluate some Gaussian functional integrals and, as it turns out, keep track of a considerable

number of indices. When all the dust clears, the self-energy of a photon of wavevector  $\mathbf{k}$  and polarization  $\lambda$  is given to lowest order in the coupling constants by

$$\Sigma(\mathbf{k}\lambda; \Omega) = i \sum_{ijlm} g_{\mathbf{k}\lambda}(ij) g_{\mathbf{k}\lambda}^{\dagger}(lm) Z_{ijlm}^{(0)}(\Omega), \quad (\text{E.8})$$

where

$$Z_{ijlm}^{(0)}(\Omega) = \int d\tau e^{i\Omega\tau} \left\{ G_{jl}^{(0)}(\tau) G_{mi}^{(0)}(-\tau) \langle \mathbf{T}[F_{ij}(\tau) F_{lm}(0)] \rangle - G_{ji}^{(0)}(\tau=0) G_{mi}^{(0)}(\tau=0) [\langle \mathbf{T}[F_{ij}(\tau) F_{lm}(0)] \rangle - \langle F_{ij} \rangle \langle F_{lm} \rangle] \right\}. \quad (\text{E.9})$$

To obtain the absorption cross-section, we consider the photon decay rate--which is given by the imaginary part of the self-energy--and multiply by  $V/c$ , where  $V$  is the volume of the world; since the coupling constants  $g \approx V^{-1/2}$ , this gives a finite result.

Thus far we have considered only the radiative interactions as a perturbation to the zero-order Hamiltonian. If we include the radiationless transitions things in principle become much more complex, but fortunately there are some simple approximations. We saw in Sect. III.A.3 that, if frequency shifts are zero, then the vertex part for the electrons is small and the generating functional for the electronic Green's functions remains Gaussian. In this case, we simply replace the zero-order electronic generating functional in Eq. (E.3) by the exact generating functional subject to the radiationless interactions, and the derivation proceeds as before; the result is that the zero-order Green's functions in  $Z$  are replaced by exact (radiationless) Green's functions.

Equation (E.8) predicts that for a single vibrational mode the absorption spectrum will consist of a sequence of lines spaced by the vibrational frequency. If we include the radiationless self-energy of the electronic Green's functions, then *if* we are in the "slow regime" each of these vibronic lines simply acquires an excess natural linewidth equal to the electronic state decay rate. On the other hand, if the electronic

dynamics are in the extreme coherent regime, the simple pattern of vibronic lines will be further split by the radiationless matrix elements, since the poles of the electronic Green's functions are displaced from their zero-order values by real rather than imaginary quantities. Of course the whole range of intermediate situations exists, and inhomogeneous broadening will blur the finer splittings, so that in general we expect the spectrum to consist of lines which are *almost*, but not quite, equally spaced.

The deviation from equal spacing is especially significant when we consider not electronic absorption but vibrational absorption, as in Sect. III.B.2. Here the time scale for relaxation of the (high-frequency) mode probed by the infrared radiation (analogous to the electronic degree of freedom in this discussion) is expected to be picoseconds, while that of the low frequency mode to which it anharmonically coupled is of order nanoseconds. While this may be an unusual situation, we never expect the relaxation of low-frequency modes to be much faster than that of high-frequency modes, so that the anharmonic coupling problem should always reside in at least the partially coherent regime where the absorption spectrum is not quite the simple pattern of vibronic lines predicted by the elementary theory.

## Notes and references, Chapter I

- 
- 1 E. Schrödinger, *What is Life?* (Cambridge University Press, Cambridge, 1944), pp. 73-74.
  - 2 For historical comments, see Volkenstein's semi-popular account, Ref. 3.
  - 3 M.V. Volkenstein, *Physics and Biology* (Academic Press, New York, 1982).
  - 4 M.A.B. Brazier, *Handbook of Physiology*, 1 (1959).
  - 5 J.C. Maxwell, *A Treatise on Electricity and Magnetism* (Clarendon Press, Oxford, 1873).
  - 6 J.P.C. Southall, editor, *Helmholtz's 'Treatise on Physiological Optics'* (Dover, New York, 1962).
  - 7 H. Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik (On the Sensation of Tone as a Physiological Basis for the Theory of Music)* (Verlag von Fr. Vieweg u. Sohn (English Edition--Dover), Braunschweig (English Edition--New York), 1862 (1954)).
  - 8 L. Pauling and E.B. Wilson, *Introduction to Quantum Mechanics* (McGraw-Hill, New York, 1935).
  - 9 L. Pauling, *Nature (Lond.)*, 161, 707-709 (1948).
  - 10 L. Pauling, D.H. Campbell, and D. Pressman, *Physiol. Rev.*, 23, 203-219 (1943).
  - 11 Lorentz' involvement in biological issues is recounted by Bouman, in Ref. 12.
  - 12 M.A. Bouman, in *Sensory Communication*, edited by W.A. Rosenblith (M.I.T press and John Wiley, Cambridge and New York, 1961), pp. 377-401.
  - 13 N. Bohr, *Nature*, 421-423 & 457-459 (1933).
  - 14 P. Jordan, *Phys. Zeits.*, 39, 711-714 (1938).
  - 15 H. Frölich, *La Revisita del Nuovo Cimento*, 7, 399-418 (1977).
  - 16 D. DeVault, *Quart. Rev. Biophys.*, 13, 387-564 (1980).
  - 17 H. Jehle, *Proc. Nat. Acad. Sci. (U.S.A.)*, 36, 238-246 (1950).
  - 18 H. Jehle, *J. Chem. Phys.*, 18, 1150-1164 (1950).
  - 19 L. Pauling and M. Delbrück, *Science*, 92, 77-79 (1940).
  - 20 J.D. Watson, *Molecular Biology of the Gene* (Benjamin, Menlo Park, California, 1976).
  - 21 L. Stryer, *Biochemistry* (W.H. Freeman, San Francisco, 1975).
  - 22 S. Sprang, D. Yang, and R.J. Fletterick, *Nature (Lond.)*, 280, 333-335 (1979).
  - 23 Some general references are 24-26.
  - 24 R.K. Clayton and W.R. Sistrom, editors, *The photosynthetic bacteria* (Plenum Press, New York, 1978).
  - 25 Govindjee, editor, *Bioenergetics of Photosynthesis* (Academic Press, New York, 1975).
  - 26 J. Barber, editor, *Primary Processes in Photosynthesis* (Elsevier, Amsterdam, 1977).
  - 27 J. Franck and E. Teller, *J. Chem. Phys.*, 6, 861-872 (1938).
  - 28 J.R. Oppenheimer, *Phys. Rev.*, 60, 158 (1941).
  - 29 D. DeVault and B. Chance, *Biophys. J.*, 6, 825-847 (1966).
  - 30 D. DeVault, J.H. Parkes, and B. Chance, *Nature*, 215, 642-644 (1967).
  - 31 T. Takano and R.E. Dickerson, *Proc. Natl. Acad. Sci. USA*, 77, 6371-6375 (1980).
  - 32 F.R. Salemme, *Ann. Rev. Biochem.*, 46, 299-329 (1977).

- 33 B. Chance, D. DeVault, H. Frauenfelder, R.A. Marcus, J.R. Schrieffer, and N. Sutin, editors, *Tunneling in Biological Systems* (Academic Press, New York, 1979).
- 34 R.S. Knox, in *Primary processes in photosynthesis, volume 2*, edited by J. Barber (Elsevier, Amsterdam New York Oxford, 1977), pp. 55-98.
- 35 E.H. Lieb, *Revs. Mod. Phys.*, *48*, 553-569 (1980).
- 36 H.F. Judson, *The Eighth Day of Creation* (Simon and Schuster, New York, 1979).
- 37 A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
- 38 A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1975).
- 39 R.P. Feynman, *Statistical Mechanics* (W.A. Benjamin, Inc., Reading, MA, 1972).
- 40 V.B. Braginsky and A.B. Manukin, *Measurement of Weak Forces in Physics Experiments*, edited by D.H. Douglass (The University of Chicago Press, Chicago and London, 1977).
- 41 C.M. Caves, K.S. Thorne, R.W.P. Drever, V.D. Sandberg, and M. Zimmerman, *Revs. Mod. Phys.*, *52*, 341-392 (1980).
- 42 C.M. Caves, *Phys. Rev. D*, *26*, 1817-1839 (1982).
- 43 M.N.G. James, *Can. J. Biochem.*, *58*, 251-271 (1980).
- 44 M. Karplus and J.A. McCammon, *C.R.C. Crit. Rev. Biochem.*, *9*, 293-349 (1981).
- 45 P.G. Debrunner and H. Frauenfelder, *Ann. Rev. Phys. Chem.*, *33*, 283-299 (1982).
- 46 D.A. Baylor, T.D. Lamb, and K.-W. Yau, *J. Physiol. (Lond.)*, *288*, 613-634 (1979).
- 47 A.J. Hudspeth and D.P. Corey, *Proc. Nat. Acad. Sci. (U.S.A.)*, *74*, 2407-2411 (1977).
- 48 R. Fettiplace and A.C. Crawford, *Proc. Roy. Soc. Lond., B.203*, 209-218 (1978).
- 49 R.H. Dicke, *Phys. Rev.*, *93*, 99-110 (1954).

## Notes and references, Chapter II

- 
- 1 E. Schrödinger, *What is Life?* (Cambridge University Press, Cambridge, 1944), p. 91.
  - 2 Lorentz' involvement with biological issues is recounted by Bouman, in Ref. 3.
  - 3 M.A. Bouman, in *Sensory Communication*, edited by W.A. Rosenblith (M.I.T press and John Wiley, Cambridge and New York, 1961), pp. 377-401.
  - 4 Hl. de Vries, *Physica*, *10*, 553-564 (1943).
  - 5 A. Rose, *J. Opt. Soc. Am.*, *38*, 196-208 (1948).
  - 6 H.B. Barlow, Cold Spring Harbor Symp. Quant. Biol., *30*, 539-546 (1965).
  - 7 S. Hecht, S. Shlaer, and M.H. Pirenne, *J. Gen. Physiol.*, *25*, 819-840 (1942).
  - 8 H.A. van der Velden, *Physica*, *11*, 179-189 (1944).
  - 9 This statement reflects a bit of hindsight. At the time of the original experiments the variations among observers caused some consternation; in particular the consistent finding of  $n = 2$  by some investigators led to specific hypotheses based on coincidence detection. More recent experiments, in which the reliability of the observer is controlled, demonstrate clearly that the apparent value of  $n$  depends on this reliability (cf. Ref. 11).
  - 10 H.F. Judson, *The Eighth Day of Creation* (Simon and Schuster, New York, 1979).
  - 11 M.C. Teich, P.R. Prucnal, G. Vannucci, M.E. Breton, and W.J. McGill, *J. Opt. Soc. Am.*, *72*, 419-431 (1982).
  - 12 M.C. Teich, P.R. Prucnal, G. Vannucci, M.E. Breton, and W.J. McGill, *Biol. Cybern.*, *44*, 157-165 (1982).
  - 13 D.A. Baylor, T.D. Lamb, and K.-W. Yau, *J. Physiol. (Lond.)*, *288*, 613-634 (1979).
  - 14 D.A. Baylor, G. Matthews, and K.-W. Yau, *J. Physiol. (Lond.)*, *309*, 591-621 (1980).
  - 15 J.L. Lawson and G.E. Uhlenbeck, *Threshold Signals* (McGraw-Hill, New York, 1950).
  - 16 D.M. Green and J.A. Swets, *Signal Detection Theory in Psychophysics* (Krieger, New York, 1966).
  - 17 R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
  - 18 There is also a non-Gaussian component which can be assigned to spontaneous photon-like events, as discussed in Appendix B and Ref. 14.
  - 19 D.A. Baylor, T.D. Lamb, and K.-W. Yau, *J. Physiol. (Lond.)*, *288*, 589-611 (1979).
  - 20 Current Topics in Membranes and Transport, *15* (1981).
  - 21 M. Rodbell, *Nature (Lond.)*, *284*, 17-22 (1979).
  - 22 B.E.A. Saleh and M.C. Teich, *Proc. I.E.E.E.*, *70*, 229-245 (1982).
  - 23 This argument for the existence and magnitude of zero-point motion works perfectly, since the numbers are the same as obtained from a rigorous calculation. This situation also obtains if we use the same argument to balance potential and (uncertainty principle induced) kinetic energies in the hydrogen atom, and one obtains the Bohr radius and ground-state ionization energy. In the oscillator case, the argument works exactly because the wave function of the ground state is Gaussian in position space, and hence in momentum space as well, so that when we average the energy we get exactly two terms proportional to each of the variances. In the hydrogen atom one has to be more judicious in

choosing factors of two and the like, although there may be some explanation of why the argument works in that case as well.

- 24 L.D. Landau and E.M. Lifshitz, *Statistical Physics* (Pergamon Press, Oxford, 1977).
- 25 R.H. Koch, Thesis, University of California, Berkeley (1982).
- 26 R.H. Koch, D.J. van Harlingen, and J. Clarke, *Phys. Rev. B.*, *26*, 74-87 (1982).
- 27 C.M. Caves, K.S. Thorne, R.W.P. Drever, V.D. Sandberg, and M. Zimmerman, *Revs. Mod. Phys.*, *52*, 341-392 (1980).
- 28 V.B. Braginsky, Y.I. Yorontsov, and K.S. Thorne, *Science*, *209*, 547-557 (1980).
- 29 E.R. Lewis, E.L. Leverenz, and W.S. Bialek, *The vertebrate inner ear* (C.R.C. Press).  
in press
- 30 W. Heiligenberg, *Principles of electrolocation and jamming avoidance in electric fish*. (Springer, Berlin, 1977).
- 31 T.H. Bullock, *Ann. Rev. Neurosci.*, *5*, 121-170 (1982).
- 32 Unpublished behavioral experiments on this point are cited by Bullock in Ref. 31.
- 33 E.I. Knudsen, *J. Comp. Physiol.*, *91*, 333-353 (1974).
- 34 G. Langner and H. Scheich, *J. Comp. Physiol.*, *128*, 235-240 (1978).
- 35 J.L. Larimer and J.A. MacDonald, *Am. J. Physiol.*, *214*, 1253-1261 (1968).
- 36 C.D. Hopkins, *J. Comp. Physiol.*, *111*, 171-207 (1976).
- 37 T.A. Viancour, *J. Comp. Physiol.*, *133*, 317-325 (1979).
- 38 S. Khanna and C. Sherrick, in *Vestibular Function and Morphology*, edited by T. Gualtierotti (Springer-Verlag, New York, Heidelberg, Berlin, 1981).
- 39 W. Bialek, unpublished calculations.
- 40 C.M. Caves, *Phys. Rev. D*, *26*, 1817-1839 (1982).
- 41 Many of the early contributors to this discussion presented their results at a symposium in 1960, the proceedings being published as Ref. 42.
- 42 C.H. Townes, editor, *Quantum Electronics: A Symposium* (Columbia University Press, New York, 1960).
- 43 H. Heffner, *Proc. I.R.E.*, *50*, 1604-1608 (1962).
- 44 The other simple case is when  $M_{11} = M_{22} = 0$ , in which case phase shifts at the input result in phase shifts at the output with opposite sign; these are phase-conjugating amplifiers. See the discussion by Caves in Ref. 40 for details.
- 45 C. Kittel, *Elementary Statistical Physics* (John Wiley and Sons, New York, 1958).
- 46 I.R. Senitzky, *Phys. Rev.*, *119*, 670-679 (1960).
- 47 H. Haken, in *Encyclopedia of Physics*, edited by S. Flugge (Springer Verlag, 1970), Vol. 25/2c.
- 48 G. von Békésy, *Experiments in Hearing* (McGraw-Hill, New York, 1960).
- 49 H. Autrum, *J. Comp. Physiol.*, *120*, 87-100 (1977).
- 50 L. Naftalin, *Cold Spring Harbor Symp. Quant. Biol.*, *30*, 169-180 (1965).
- 51 W.S. Rhode, *J. Acoust. Soc. Am.*, *64*, 158-176 (1978).
- 52 A.J. Hudspeth and R. Jacobs, *Proc. Nat. Acad. Sci. (U.S.A.)*, *76*, 1506-1509 (1979).
- 53 L.D. Landau and E.M. Lifshitz, *Fluid Mechanics* (Pergamon Press, Oxford, 1959).

- 54 G. Zweig, Cold Spring Harbor Symp. Quant. Biol., 40, 619-633 (1975).
- 55 H. Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik (On the Sensation of Tone as a Physiological Basis for the Theory of Music)* (Verlag von Fr. Vieweg u. Sohn (English Edition--Dover), Braunschweig (English Edition--New York), 1862 (1954)).
- 56 J. Zwislocki, J. Acoust. Soc. Am., 25, 743-751 (1953).
- 57 O.F. Ranke, Zeits. f. Biol., 103, 409-434 (1950).
- 58 L.C. Peterson and B.P. Bogert, J. Acoust. Soc. Am., 22, 369-381 (1950).
- 59 I.J. Russell and P.M. Sellick, J. Physiol. (Lond.), 284, 261-290 (1978).
- 60 E.F. Evans, Handbk. Sens. Physiol., V/2, 1-108 (1975).
- 61 E.F. Evans and J.P. Wilson, Science, 190, 1218-1221 (1975).
- 62 P. Hillman, H. Schechter, and M. Rubenstein, Rev. Mod. Phys., 36, 360 (1964).
- 63 P. Gilad, H. Schechter, P. Hillman, and M. Rubenstein, J. Acoust. Soc. Am., 41, 1232-1236 (1967).
- 64 B.M. Johnstone and A.J.F. Boyle, Science, 158, 389-390 (1967).
- 65 W.S. Rhode, J. Acoust. Soc. Am., 49, 1218-1231 (1971).
- 66 W.S. Rhode, J. Acoust. Soc. Am., 67, 1696-1703 (1980).
- 67 P.M. Sellick, R. Patuzzi, and B.M. Johnstone, J. Acoust. Soc. Am., 72, 131-141 (1982).
- 68 J.P. Wilson and J.R. Johnstone, J. Acoust. Soc. Am., 57, 705-723 (1975).
- 69 E.L. LePage and B.M. Johnstone, Hearing Res., 2, 183-189 (1980).
- 70 L.U.E. Kohlöffel, Acoustica, 27, 49-65 (1972).
- 71 S.M. Khanna and D.G.B. Leonard, Science, 215, 305-306 (1982).
- 72 L.U.E. Kohlöffel, Acoustica, 27, 66-81 (1972).
- 73 L.U.E. Kohlöffel, Acoustica, 27, 82-89 (1972).
- 74 Khanna and Leonard (Ref. 71) make clear that none of the the animals which they studied are unaffected by the measurement procedure. While the measurement is taking place, it is possible to observe resonances in the input impedance of the ear which were not present before the cochlea was opened and the mirror (for interferometry) dropped onto the membrane. After the experiment is over, the cochleae are examined under the microscope, and it is clear that there has been significant structural damage in the neighborhood of the mirror, apparently resulting from the immune reaction which the mirror elicits in the animal. This is in spite of the use of single-crystal  $10^{-8}$  gm gold mirrors, which are alleged to be chemically inert.
- 75 To determine the threshold sound pressure *at the eardrum* one must use the behavioral data on free-field thresholds together with acoustical measurements on the sound pressure transformation from the free-field to the eardrum; such data have been gathered with considerable care for the cat, as described in the primary references, 76 and 77 below. The calculation of eardrum thresholds has been reported by Lynch *et al.* in Ref. 78.
- 76 J.D. Miller, C.S. Watson, and W.P. Covell, Acta Otolaryngol., 5176, 1-87 (1963).
- 77 F.M. Wiener, R.R. Pfeifer, and A.S.N. Backus, Acta Otolaryngol., 61, 255-269 (1965).
- 78 T.J. Lynch, III, V. Nedzelnitsky, and W.T. Peake, J. Acoust. Soc. Am., 72, 108-130 (1982).

- 79 R. Fettiplace and A.C. Crawford, Proc. Roy. Soc. Lond., *B.203*, 209-218 (1978).
- 80 D.A. Goodman, R.L. Smith, and S.C. Chamberlain, Hearing Res., *7*, 161-179 (1982).
- 81 D.H. Johnson, Thesis, Massachusetts Institute of Technology (1974).
- 82 D.H. Johnson, J. Acoust. Soc. Am., *68*, 1115-1122 (1980).
- 83 V. Nedzelnsky, Thesis, Massachusetts Institute of Technology (1974).
- 84 V. Nedzelnsky, J. Acoust. Soc. Am., *68*, 1676-1689 (1980).
- 85 W.T. Peake and A.R. Ling, Jr, J. Acoust. Soc. Am., *67*, 1736-1745 (1980).
- 86 T.F. Weiss, W.T. Peake, A. Ling, Jr, and T. Holton, in *Evoked Electrical Activity in the Auditory Nervous System*, edited by R. Nauton (Academic, New York, 1978), pp. 91-112.
- 87 T. Holton and T.F. Weiss, Brain Res., *159*, 219-222 (1978).
- 88 A.C. Crawford and R. Fettiplace, J. Physiol. (Lond.), *306*, 79-125 (1980). E.R. Lewis and co-workers, unpublished..ds [F 90
- 89 D.J. Lim, J. Acoust. Soc. Am., *67*, 1686-1695 (1980).
- 91 L.G. Tilney and J.C. Saunders, J. Cell Biol., *96*, 807-821 (1983).
- 92 A. Flock and H.C. Cheung, J. Cell. Biol., *75*, 339-343 (1977).
- 93 M. Karplus and J.A. McCammon, C.R.C. Crit. Rev. Biochem., *9*, 293-349 (1981). E.R. Lewis and co-workers, unpublished..ds [F 95
- 94 A.J. Hudspeth and D.P. Corey, Proc. Nat. Acad. Sci. (U.S.A.), *74*, 2407-2411 (1977).
- 96 J. Lighthill, *Mathematical Biofluidynamics* (SIAM, Philadelphia, 1975).
- 97 M.N. Flerov, Biofizika, *21*, 1092-1096 (1976).
- 98 E. de Boer, Phys. Repts. (Phys. Lett. C), *62*, 87-174 (1980).
- 99 D.C.S. White and J. Thorson, Prog. Biophys. Mol. Biol., *27*, 175-255 (1973).
- 100 D.C.S. White, M.G.A. Wilson, and J. Thorson, in *Cross-bridge mechanism in muscle contraction*, edited by H. Sugi and G.H. Pollack (University Park Press, Baltimore, 1979), pp. 194-209.
- 101 M.G.A. Wilson, Thesis, University of York (1979).
- 102 M.G.A. Wilson and D.C.S. White, J. Muscle Res. Cell Motility, *4*, 283-306 (1983).
- 103 J.C. Macartney, S.D. Comis, and J.O. Pickles, Nature, *288*, 491-492 (1980).
- 104 N. Hirokawa and L.G. Tilney, J. Cell. Biol., *95*, 249-261 (1982).
- 105 A. Flock, H.C. Cheung, B. Flock, and G. Utter, J. Neurocytol., *10*, 133-147 (1981).
- 106 K.S. Cole, *Membranes, Ions, and Impulses* (University of California Press, Berkeley, 1972).
- 107 A.L. Hodgkin and A.F. Huxley, J. Physiol. (Lond.), *117*, 500-544 (1952).
- 108 J.J.B. Jack, D. Noble, and R.W. Tsien, *Electric Current Flow in Excitable Cells* (Clarendon Press, Oxford, 1975).
- 109 W.T. Clusin and M.V.L. Bennett, J. Gen. Physiol., *73*, 703-723 (1979).
- 110 T.A. Viancour, J. Comp. Physiol., *133*, 327-338 (1979).
- 111 D.P. Corey and A.J. Hudspeth, Biophys. J., *26*, 499-506 (1979).
- 112 Hl. de Vries, Nature (Lond.), *161*, 63 (1948). E.R. Lewis and co-workers, unpublished..ds [F 114

- 113 P.L. Richards, T.M. Shen, R.E. Harris, and F.L. Lloyd, *Appl. Phys. Lett.*, *36*, 480-482 (1980).
- 115 J. Weber, *Phys. Rev.*, *108*, 537-541 (1957). E.R. Lewis and co-workers, unpublished.  
[F 117]
- 116 W.H. Louisell, A. Yariv, and A.E. Siegman, *Phys. Rev.*, *124*, 1646-1654 (1961).

## Notes and references, Chapter III

- 
- 1 P.A.M. Dirac, *Principles of Quantum Mechanics* (Oxford University Press, Oxford, 1967 (revised 4<sup>th</sup> edition; 1<sup>st</sup> edition 1930)).
  - 2 D. DeVault and B. Chance, *Biophys. J.*, **6**, 825-847 (1966).
  - 3 D. DeVault, J.H. Parkes, and B. Chance, *Nature*, **215**, 642-644 (1967).
  - 4 V.I. Goldanskii, *Ann. Rev. Phys. Chem.*, **27**, 85-126 (1976).
  - 5 L.I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968).
  - 6 H. Eyring, *Chem. Rev.*, **17**, 65-77 (1935).
  - 7 F.R. Salemme, *Ann. Rev. Biochem.*, **46**, 299-329 (1977).
  - 8 L.N. Grigorov and D.S. Chernavskii, *Biofizika*, **17**, 195-202 (1972).
  - 9 J.J. Hopfield, *Proc. Nat. Acad. Sci. USA*, **71**, 3640-3644 (1974).
  - 10 E.U. Condon, *Phys. Rev.*, **32**, 858-872 (1928).
  - 11 E.E. Nikitin, *Theory of Elementary Atomic and Molecular Processes in Gases*, edited by M.J. Kearsley (Clarendon Press, Oxford, 1974).
  - 12 L. Pauling, *Nature (Lond.)*, **161**, 707-709 (1948).
  - 13 A. Fersht, *Enzyme Structure and Function* (Freeman, San Francisco, 1977).
  - 14 W.P. Jencks, *Catalysis in Chemistry and Enzymology* (McGraw-Hill, New York, 1969).
  - 15 J. Jortner, *J. Chem. Phys.*, **64**, 4860-4867 (1976).
  - 16 A.M. Stoneham, *Rep. Prog. Phys.*, **44**, 1251-1295 (1981).
  - 17 K. Huang, *Contemp. Phys.*, **22**, 599-612 (1981).
  - 18 T.F. Soules and C.B. Duke, *Phys. Rev. B*, **3**, 262-274 (1971).
  - 19 H. Haken and G. Strobl, *Z. Physik*, **262**, 135-148 (1973).
  - 20 R. Wertheimer and R. Silbey, *Chem. Phys. Lett.*, **75**, 243-248 (1980).
  - 21 R.W. Munn and R. Silbey, *J. Chem. Phys.*, **68**, 2439-2450 (1978).
  - 22 S. Rackovsky and R. Silbey, *Mol. Phys.*, **25**, 61-72 (1973).
  - 23 I.I. Abram and R. Silbey, *J. Chem. Phys.*, **63**, 2317-2328 (1975).
  - 24 V.M. Kenkre and R.S. Knox, *Phys. Rev. B*, **9**, 5279-5290 (1974).
  - 25 T.S. Rahman, R. S. Knox, and V.M. Kenkre, *Chem. Phys.*, **44**, 197-211 (1979).
  - 26 L.J. Noe, W.G. Eisert, and P.M. Rentzepis, *Proc. Nat. Acad. Sci. (U.S.A.)*, **75**, 573-577 (1978).
  - 27 R.H. Austin, K.W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus, *Biochemistry*, **14**, 5355-5373 (1975).
  - 28 N. Alberding, K.W. Beeson, S.S. Chan, L. Eisenstein, H. Frauenfelder, and T.M. Nordlund, *Science*, **192**, 1002-1004 (1976).
  - 29 N. Alberding, R.H. Austin, S.S. Chan, L. Eisenstein, H. Frauenfelder, I.C. Gunsalus, and T.M. Nordlund, *J. Chem. Phys.*, **65**, 4701-4711 (1976).
  - 30 N. Alberding, S.S. Chan, L. Eisenstein, H. Frauenfelder, D. Good, I.C. Gunsalus, T.M. Nordlund, M.F. Perutz, A.H. Reynolds, and L.B. Sorenson, *Biochemistry*, **17**, 43-51 (1978).
  - 31 H. Frauenfelder, in *Tunneling in Biological Systems*, edited by B. Chance, D. DeVault, H. Frauenfelder, and R. Marcus (Academic, New York, 1979), pp. 624-649.

- 32 D. Kleinfeld, M.Y. Okamura, and G. Feher, *Biophys. J.*, *41*, 121a (1983).
- 33 D. Holtén and M.W. Windsor, *Ann. Rev. Biophys. Bioeng.*, *7*, 189-227 (1978).
- 34 E.F. Hilinski and P. Rentzepis, *Nature (Lond.)*, *302*, 481-487 (1983).
- 35 R.K. Clayton and W.R. Sistrom, editors, *The photosynthetic bacteria* (Plenum Press, New York, 1978).
- 36 M.Y. Okamura, R.A. Isaacson, and G. Feher, *Biochim. Biophys. Acta*, *546*, 394-417 (1979).
- 37 K. Peters, Ph. Avouris, and P.M. Rentzepis, *Biophys. J.*, *23*, 207-217 (1978).
- 38 C.C. Schenck, W.W. Parson, D. Holtén, M.W. Windsor, and A. Sarai, *Biophys. J.*, *36*, 479-489 (1981).
- 39 J. Jortner, *Biochim. Biophys. Acta*, *594*, 193-230 (1980).
- 40 K.K. Rebane and R.A. Avarmaa, *Chem. Phys.*, *68*, 191-200 (1982).
- 41 J. Randall and J.M. Vaughn, *Philos. Trans. Roy. Soc. Lond.*, *A293*, 341-347 (1979).
- 42 I.R. Senitzky, *Phys. Rev.*, *119*, 670-679 (1960).
- 43 I.R. Senitzky, *Phys. Rev.*, *131*, 2827-2838 (1963).
- 44 M. Lax, *Phys. Rev.*, *129*, 2342-2348 (1963).
- 45 H. Haken, in *Encyclopedia of Physics*, edited by S. Flugge (Springer Verlag, 1970), Vol. 25/2c.
- 46 Note that if  $\delta F$  were a c-number rather than an operator, the commutation relations would not be preserved. This would correspond to an (incorrect) picture in which coupling to the heat bath could dissipate all of the zero-point energy, leaving the system with no quantum noise. Haken (Ref. 45) has emphasized the view of the "quantum Langevin" force as having precisely the operator properties required to avoid this inconsistency with quantum mechanics.
- 47 R.M. Levy and M. Karplus, *Biopolymers*, *18*, 2465-2495 (1979).
- 48 A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
- 49 For simplicity all of these introductory calculations will be done at zero temperature, so that  $\langle \dots \rangle$  denotes a ground state average (vacuum expectation value) rather than a thermal average. Extensions to finite temperature and to averages over non-equilibrium ensembles will be discussed below.
- 50 In more complex model Hamiltonians the imaginary part will determine the *dephasing* rate analogous to  $T_2^{-1}$  in magnetic systems, and we will have to examine higher order Green's functions to distinguish the lifetime broadening and "pure dephasing" contributions to  $T_2$ , as discussed in more detail below.
- 51 R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
- 52 C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
- 53 P. Ramond, *Field Theory: A Modern Primer* (Benjamin/Cummings, Reading, MA, 1981).
- 54 R.F. Fox, *Phys. Repts.*, *48*, 179-283 (1978).
- 55 I assume the the averages of the  $M_{ij}$  are zero. The breakdown of this approximation is relevant only if two of the electronic states are near-degenerate or if we consider interactions with the radiation field; these two case are discussed in Appendices D and E, respectively.
- 56 A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1975).

- 57 A.G. Redfield, IBM J. Research Develop., *1*, 19-31 (1957).
- 58 R.F. Goldstein and W. Bialek, Phys. Rev. B, *27*, 7431-7439 (1983).
- 59 K.A. Brueckner and C.A. Levinson, Phys. Rev., *97*, 1344-1352 (1955).
- 60 M. Gell-Mann and K.A. Brueckner, Phys. Rev., *106*, 364-368 (1957).
- 61 Recall that in the absence of pure dephasing terms the irreducible part of the two-particle Green's function, which is the first non-Gaussian term in the generating functional, is perturbatively small.
- 62 R.F. Goldstein, personal communication.
- 63 R.F. Goldstein, Thesis, University of California, Berkeley (1982).
- 64 T. Takano and R.E. Dickerson, Proc. Natl. Acad. Sci. USA, *77*, 6371-6375 (1980).
- 65 J.J. Hopfield, Biophys. J., *18*, 311-321 (1977).
- 66 R.F. Goldstein and A.J. Bearden, Proc. Nat. Acad. Sci. (U.S.A.)  
in the press.
- 67 T.C. Streckas and T.G. Spiro, Biochim. Biophys. Acta, *278*, 188-192 (1972).
- 68 J.M. Friedman, D.L. Rousseau, and F. Adar, Proc. Natl. Acad. Sci. USA, *74*, 2607-2611 (1977).
- 69 A. Laubereau, Philos. Trans. Roy. Soc. Lond., *A293*, 441-454 (1979).
- 70 P. Jordan, Phys. Zeits., *39*, 711-714 (1938).
- 71 H. Jehle, Proc. Nat. Acad. Sci. (U.S.A.), *36*, 238-246 (1950).
- 72 H. Jehle, J. Chem. Phys., *18*, 1150-1164 (1950).
- 73 L. Pauling and M. Delbrück, Science, *92*, 77-79 (1940).
- 74 This assumes that the frequency of any particular molecule in the sample, while chosen out of the distribution  $P(\omega)$ , does not change on the time scale of the experiment; this point will be discussed below for a concrete example.
- 75 M. Tsubaki, R.B. Srivastava, and N.-T. Yu, Biochemistry, *21*, 1132-1140 (1982).
- 76 L. Powers, Biochim. Biophys. Acta, *683*, 1-38 (1982).
- 77 J.O. Alben, D. Beece, S.F. Bowne, W. Doster, L. Eisenstein, H. Frauenfelder, D. Good, J.D. McDonald, M.C. Marden, P.P. Moh, L. Reinisch, A.H. Reynolds, E. Shyamsunder, and K.T. Yue, Proc. Nat. Acad. Sci. (U.S.A.), *79*, 3744-3748 (1982).
- 78 T. Kitagawa, K. Nagai, and M. Tsubaki, FEBS Letters, *104*, 376-378 (1979).
- 79 S.E.V. Phillips, J. Mol. Biol., *142*, 531-554 (1980).
- 80 M.H. Redi, B.S. Gerstman, and J.J. Hopfield, Biophys. J., *35*, 471-484 (1981).
- 81 J.O. Alben, D. Beece, S.F. Bowne, L. Eisenstein, H. Frauenfelder, D. Good, M.C. Marden, P.P. Moh, L. Reinisch, A.H. Reynolds, and K.T. Yue, Phys. Rev. Lett., *44*, 1157-1160 (1980).
- 82 J.C. Hanson and B.P. Schoenborn, J. Mol. Biol., *153*, 117-146 (1981).
- 83 J.M. Ziman, *Models of Disorder* (Cambridge Univ. Press, Cambridge, 1979).
- 84 P.M. Morse, Phys. Rev., *34*, 57-64 (1929).
- 85 K.D. Watenpugh, T.N. Margulis, L.C. Sieker, and L.H. Jensen, J. Mol. Biol., *122*, 175-190 (1978).
- 86 A.R. Sielecki, W.A. Hendrickson, C.G. Broughton, L.T.J. Delabaere, G.D. Brayer, and M.N.G. James, J. Mol. Biol., *134*, 781-804 (1979).
- 87 C.C.F. Blake, W.C.A. Pulford, and P.J. Artymiuk, J. Mol. Biol., *167*, 693-723 (1983).
- 88 H.D. Middendorf and J. Randall, Philos. Trans. Roy. Soc. Lond., *B290*, 639-854 (1980).

- 89 L. Eisenstein and H. Frauenfelder, in *Frontiers of Biological Energetics. I: Electrons to Tissues*, edited by P.L. Dutton, J.S. Leigh, and A. Scarpa (Academic Press, New York, 1979), pp. 680-688.
- 90 M.W. Makinen, R.A. Houtchens, and W.S. Caughey, *Proc. Nat. Acad. Sci. U.S.A.*, **76**, 6042-6046 (1979).
- 91 W.S. Caughey, H. Shimada, M.G. Choc, and M.P. Tucker, *Proc. Nat. Acad. Sci. (U.S.A.)*, **78**, 2903-2907 (1981).
- 92 C. Kittel, *Introduction to Solid State Physics* (John Wiley, New York, 1971).
- 93 H. Frauenfelder, *The Mössbauer Effect* (W.A. Benjamin, New York, 1962).
- 94 H. Frauenfelder, G.A. Petsko, and D. Tsernoglou, *Nature (Lond.)*, **280**, 558-563 (1979).
- 95 H. Hartman, F. Parak, W. Steigemann, G.A. Petsko, D. Ringe Ponzi, and H. Frauenfelder, *Proc. Nat. Acad. Sci. (U.S.A.)*, **79**, 4967-4971 (1982).
- 96 H. Keller and P. Debrunner, *Phys. Rev. Lett.*, **45**, 68-71 (1980).
- 97 F. Parak, E.N. Frolov, R.L. Mössbauer, and V.I. Goldanskii, *J. Mol. Biol.*, **145**, 825-833 (1981).
- 98 F. Parak, E.W. Knapp, and D. Kucheida, *J. Mol. Biol.*, **161**, 177-194 (1982).
- 99 E.R. Bauminger, S.G. Cohen, I. Nowick, S. Ofer, and J. Yariv, *Proc. Nat. Acad. Sci. (U.S.A.)*, **80**, 736-740 (1983).
- 100 P.G. Debrunner and H. Frauenfelder, *Ann. Rev. Phys. Chem.*, **33**, 283-299 (1982).
- 101 C. Kittel, *Quantum Theory of Solids* (J. Wiley & Sons, New York, 1963), pp. 148-149.
- 102 T. Miyazawa, *J. Mol. Spectrosc.*, **13**, 321-325 (1964).
- 103 A. Desbois, M. Lutz, and R. Banerjee, *Biochemistry*, **18**, 1510-1518 (1979).
- 104 J.M. Ziman, *Principles of the Theory of Solids* (Cambridge Univ. Press, Cambridge).
- 105 The overall Debye-Waller factor is obtained from the *average* decay of diffraction intensity at increasing scattering angle, and indicates the magnitude of motions which the entire molecule exhibits as a whole. It is thus not subject to the independent motion approximation criticized above.
- 106 S.W. Lovesey, *Condensed Matter Physics: Dynamics Correlations* (Benjamin/Cummings, Reading, MA, 1980).
- 107 W. Bialek and R.F. Goldstein, unpublished calculations.
- 108 Obviously this cannot be done without some qualification, since time and frequency are complementary variables. In practice a natural and consistent discretization is forced by the properties of the optical pulse used in the experiment.
- 109 A. Abragam, *Principles of Nuclear Magnetism* (Clarendon Press, Oxford, 1961).
- 110 A.J. Hoff, *Phys. Repts. (Phys. Lett. C)*, **54**, 75-200 (1979).
- 111 A.J. Hoff, *Quart. Rev. Biophys.*, **14**, 599-665 (1981).
- 112 P.W. Anderson, *Phys. Rev.*, **115**, 2-13 (1959).
- 113 M.Y. Okamura, D.R. Fredkin, R.A. Isaacson, and G. Feher, in *Tunneling in Biological Systems*, edited by B. Chance, D.C. DeVault, H. Frauenfelder, and R.A. Marcus (Academic Press, New York, 1979), pp. 729-743.
- 114 R. Haberkorn, M.E. Michel-Beyerle, and R.A. Marcus, *Proc. Nat. Acad. Sci. (U.S.A.)*, **76**, 4185-4188 (1979).
- 115 Rigorously we should consider eight states, taking account of the full triplet manifold. This adds much complexity and no new physics; inclusion of the full triplet manifold will be required, however, for a quantitative treatment of anisotropies in the exchange interaction (see below).

- 116 J.M. Ziman, *Elements of Advanced Quantum Theory* (Cambridge University Press, Cambridge, 1969), pp. 65-66.
- 117 R. Englman and J. Jortner, *Mol. Phys.*, *18*, 145-164 (1970).
- 118 B.D. Fried and S.D. Conte, *The Plasma Dispersion Function* (Academic Press, New York, 1961).
- 119 S.G. Boxer, C.E.D. Chidsey, and M.G. Roelofs, *Proc. Nat. Acad. Sci. (U.S.A.)*, *79*, 4632-4636 (1982).
- 120 A. Ogrodnik, H.W. Krüger, H. Orthuber, R. Haberkorn, M.E. Michel-Beyerle, and H. Scheer, *Biophys. J.*, *39*, 91-99 (1982).
- 121 J.R. Norris, M.K. Bowman, D.E. Budil, J. Tang, C.A. Wraight, and G.L. Closs, *Proc. Nat. Acad. Sci. (U.S.A.)*, *79*, 5532-5536 (1982).
- 122 D. Holton, C. Hoganson, M.W. Windsor, C.C. Schenck, W.W. Parson, A. Migus, R.L. Fork, and C.V. Shank, *Biochim. Biophys. Acta*, *592*, 461-477 (1980).
- 123 Or, more generally, light of finite bandwidth  $\Delta\Omega$  such that the pulse is not transform limited, *i.e.*  $\Delta\Omega t \gg 1$ .
- 124 D. deLeeuw, M. Malley, G. Butterman, M.Y. Okamura, and G. Feher, *Biophys. J.*, *37*, 111a (1982).
- 125 P.M. Champion and A.C. Albrecht, *Ann. Rev. Phys. Chem.*, *33*, 353-376 (1982).
- 126 *Biophys. J.*, *32*, 1-676 (1980).
- 127 R.H. Dicke, *Phys. Rev.*, *93*, 99-110 (1954).
- 128 R.J. Glauber, in *Quantum Optics and Electronics*, edited by C. DeWitt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1964), pp. 65-185.
- 129 N.A. Curtin, C. Gilbert, K.M. Kretzschmar, and D.R. Wilkie, *J. Physiol. (Lond.)*, *269*, 255- (1974).
- 130 J.D. Watson, *Molecular Biology of the Gene* (Benjamin, Menlo Park, California, 1976).
- 131 D.R. Griffin, *Listening in the Dark: The acoustic orientation of bats and men* (Dover, New York, 1974).
- 132 W.H. Louisell, A. Yariv, and A.E. Siegman, *Phys. Rev.*, *124*, 1646-1654 (1961).
- 133 C.M. Caves, *Phys. Rev. D*, *26*, 1817-1839 (1982).
- 134 J.M. Squire, *Ann. Rev. Biophys. Bioeng.*, *4*, 137-163 (1975).
- 135 A. Flock and H.C. Cheung, *J. Cell. Biol.*, *75*, 339-343 (1977).
- 136 L.G. Tilney, D.J. deRosier, and M.J. Mulroy, *J. Cell. Biol.*, *86*, 244-259 (1980).
- 137 N. Hirokawa and L.G. Tilney, *J. Cell. Biol.*, *95*, 249-261 (1982).
- 138 A. Flock, H.C. Cheung, B. Flock, and G. Utter, *J. Neurocytol.*, *10*, 133-147 (1981).
- 139 L.D. Landau and E.M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1960).
- 140 B. Hudson, L. Ziegler, D. Flamig, and D. Gerrity, *Biophys. J.*, *41*, 222a (1983).
- 141 T. Holton, *Hearing Res.*, *2*, 21-38 (1980).
- 142 E.R. Lewis and co-workers, unpublished.
- 143 D.O. Kim, *Hearing Res.*, *2*, 297-396 (1980).
- 144 G. Pollack, O.W. Henson, Jr, and A. Novick, *Science*, *176*, 66-68 (1972).
- 145 T.F. Weiss, W.T. Peake, A. Ling, Jr, and T. Holton, in *Evoked Electrical Activity in the Auditory Nervous System*, edited by R. Nauton (Academic, New York, 1978), pp. 91-112.
- 146 R.G. Turner, A.A. Muraski, and D.W. Nielsen, *Science*, *213*, 1519-1521 (1981).

- 147 L.G. Tilney and J.C. Saunders, *J. Cell Biol.*, *96*, 807-821 (1983).
- 148 D.J. Lim, *J. Acoust. Soc. Am.*, *67*, 1686-1695 (1980).
- 149 W.T. Peake and A.R. Ling, Jr, *J. Acoust. Soc. Am.*, *67*, 1736-1745 (1980).
- 150 A. Flock, B. Flock, and E. Murray, *Acta Oto-Laryngol.*, *83*, 85-91 (1977).
- 151 T.F. Weiss, *Hearing Res.*, *7*, 353-360 (1982).
- 152 T. Gold, *Proc. Roy. Soc. Edinb.*, *B135*, 492-498 (1948).
- 153 D.O. Kim, S.T. Neely, C.E. Molnar, and J.W. Matthews, in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F.A. Bilsen (Delft University Press, Delft, 1980), pp. 7-14.
- 154 V. Bruns and M. Goldbach, *Anat. Embryol.*, *161*, 51-63 (1980).
- 155 M. Born and E. Wolf, *Principles of Optics (6<sup>th</sup> edition)* (Pergamon Press, Oxford, New York, 1980).
- 156 N. Suga, G. Neuweiler, and J. Möller, *J. Comp. Physiol.*, *106*, 111-125 (1976).
- 157 L.D. Landau and E.M. Lifshitz, *Statistical Physics* (Pergamon Press, Oxford, 1977).
- 158 W. Bialek, in *Hearing: Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman (Springer-Verlag, Berlin, in the press, 1983).
- 159 D.C.S. White and J. Thorson, *Prog. Biophys. Mol. Biol.*, *27*, 175-255 (1973).
- 160 P.M. Zurek, *J. Acoust. Soc. Am.*, *69*, 514-523 (1981).
- 161 P.M. Zurek and W.W. Clark, *J. Acoust. Soc. Am.*, *70*, 446-450 (1981).
- 162 J.P. Wilson, *Hearing Res.*, *2*, 233-252 (1980).
- 163 A.R. Palmer and J.P. Wilson, *J. Physiol. (Lond.)*, *324*, 66P (1981).
- 164 D.T. Kemp, *Arch. Otorhinolaryngol.*, *224*, 37-45 (1979).
- 165 H.P. Wit, J.C. Langevoort, and R.J. Ritsma, *J. Acoust. Soc. Am.*, *70*, 437-445 (1981).
- 166 W.M. Rabinowitz and G.P. Widin, *J. Acoust. Soc. Am.*, *70*, S7 (1981).
- 167 W. Bialek and A.L. Schweitzer, unpublished.
- 168 In making this comparison it is important to note that not all the stereocilia on a single receptor cell are of the same length (cf. Ref's. 146-149). In order that these variations not destroy the filtering achieved by the feedback process, there must be some relations among the lengths of the various cilia, e.g. that they are all integer multiples of some fundamental length; this may be tested by careful anatomical studies using quantitative electron microscopy. Such variations would have the potentially useful effect of suppressing the secondary resonances associated with the individual cilia; again this point should be kept in mind in comparing theory with experiment.

## Notes and references, Chapter IV

- 
- 1 W. Heisenberg, *The Physical Principles of the Quantum Theory* (Dover, New York, 1949), p. 10.
  - 2 For a recent example of this approach, see Ref. 25
  - 3 W.P. Jencks, *Catalysis in Chemistry and Enzymology* (McGraw-Hill, New York, 1969).
  - 4 A. Fersht, *Enzyme Structure and Function* (Freeman, San Francisco, 1977).
  - 5 This approach *via* condensed matter theory was first emphasized by Hopfield. See his and other contributions in Ref. 33 of Chapter I.
  - 6 J. Bardeen and D. Pines, *Phys. Rev.*, **99**, 1140-1150 (1955).
  - 7 A.B. Migdal, *Sov. Phys. J.E.T.P.*, **34**, 1438-1446 (1958).
  - 8 R.P. Feynman, *Statistical Mechanics* (W.A. Benjamin, Inc., Reading, MA, 1972), pp. 221-241.
  - 9 T. Holstein, *Ann. Phys. (N.Y.)*, **8**, 325-342 (1959).
  - 10 A.M. Stoneham, *Rep. Prog. Phys.*, **44**, 1251-1295 (1981).
  - 11 W.P. Su, J.R. Schrieffer, and A.J. Heeger, *Phys. Rev. B.*, **22**, 2099-2111 (1980).
  - 12 J.P. Sethna, *Phys. Rev. B*, **24**, 698-713 (1981).
  - 13 A.L. Kholodenko and K.F. Freed, *Phys. Rev. B*, **27**, 4586-4600 (1983).
  - 14 Electronic coherence also persists for long times in those reactions which are slow compared to vibrational relaxation provided that certain conditions are met. These conditions are always fulfilled at low temperature, and experiments on the decay of coherence in low-temperature systems are in accord with this prediction; cf. Appendix D.
  - 15 R.H. Austin, K.W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus, *Biochemistry*, **14**, 5355-5373 (1975).
  - 16 J.O. Alben, D. Beece, S.F. Bowne, L. Eisenstein, H. Frauenfelder, D. Good, M.C. Marden, P.P. Moh, L. Reinisch, A.H. Reynolds, and K.T. Yue, *Phys. Rev. Lett.*, **44**, 1157-1160 (1980).
  - 17 D. deLeeuw, M. Malley, G. Butterman, M.Y. Okamura, and G. Feher, *Biophys. J.*, **37**, 111a (1982).
  - 18 L. Stryer, *Biochemistry* (W.H. Freeman, San Francisco, 1975).
  - 19 J.M. Sturtevant, *Proc. Nat. Acad. Sci. (U.S.A.)*, **74**, 2236-2240 (1977).
  - 20 A.R. Sielecki, W.A. Hendrickson, C.G. Broughton, L.T.J. Delabaere, G.D. Brayer, and M.N.G. James, *J. Mol. Biol.*, **134**, 781-804 (1979).
  - 21 M.N.G. James, A.R. Sielecki, G.D. Brayer, L.T.J. Delabaere, and C.A. Bauer, *J. Mol. Biol.*, **144**, 43-88 (1980).
  - 22 T. Takano and R.E. Dickerson, *Proc. Natl. Acad. Sci. USA*, **77**, 6371-6375 (1980).
  - 23 M.N.G. James, *Can. J. Biochem.*, **58**, 251-271 (1980).
  - 24 F.R. Salemme, in *Frontiers of Biological Energetics. I: Electrons to Tissues*, edited by P.L. Dutton, J.S. Leigh, and A. Scarpa (Academic Press, New York, 1978), pp. 83-90.
  - 25 C.P. Huber, Y. Ozaki, D.H. Pliura, A.C. Storer, and P.R. Carey, *Biochemistry*, **21**, 3109-3115 (1982).
  - 26 B. Hudson, L. Ziegler, D. Flamig, and D. Gerrity, *Biophys. J.*, **41**, 222a (1983).
  - 27 J.G. Belasco and J.R. Knowles, *Biochemistry*, **19**, 472-477 (1980).

- 28 T.L. Hill, *Free Energy Transduction in Biology* (Academic Press, New York, 1977).
- 29 W. Bialek, unpublished.
- 30 H. Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grunlage fur die Theorie der Musik (On the Sensation of Tone as a Physiological Basis for the Theory of Music)* (Verlag von Fr. Vieweg u. Sohn (English Edition--Dover), Braushcweig (English Edition--New York), 1862 (1954)).
- 31 G. von Békésy, *Experiments in Hearing* (McGraw-Hill, New York, 1960).
- 32 T.H. Bullock, *Ann. Rev. Neurosci.*, 5, 121-170 (1982).

## Notes and references, Appendices

- 
- 1 J.L. Lawson and G.E. Uhlenbeck, *Threshold Signals* (McGraw-Hill, New York, 1950).
  - 2 D.M. Green and J.A. Swets, *Signal Detection Theory in Psychophysics* (Krieger, New York, 1966).
  - 3 A.C. Crawford and R. Fettiplace, *J. Physiol. (Lond.)*, 306, 79-125 (1980).
  - 4 D.A. Goodman, R.L. Smith, and S.C. Chamberlain, *Hearing Res.*, 7, 161-179 (1982).
  - 5 W.M. Siebert, *Kybernetik*, 2, 206-215 (1965).
  - 6 E.F. Evans, *Handbk. Sens. Physiol.*, VI/2, 1-108 (1975).
  - 7 D.H. Johnson, Thesis, Massachusetts Institute of Technology (1974).
  - 8 D.H. Johnson, *J. Acoust. Soc. Am.*, 68, 1115-1122 (1980).
  - 9 This result is strictly applicable only to situations in which the observer has *a priori* knowledge concerning the phase of the incoming signal. In general this is not the case, and we may construct new probability distributions which reflect this phase uncertainty; unfortunately subsequent calculations with these distributions prove to be quite difficult. One expects by analogy with continuous signals, however, that complete phase uncertainty would result simply in a 3 dB decrease in signal-to-noise ratio, although I know of no proof of this result for the case of modulated Poisson processes considered here.
  - 10 J.A. Costalupes, *Hearing Res.*, 9, 43-54 (1983).
  - 11 T.J. Lynch, III, V. Nedzelitsky, and W.T. Peake, *J. Acoust. Soc. Am.*, 72, 108-130 (1982).
  - 12 J.D. Miller, C.S. Watson, and W.P. Covell, *Acta Otolaryngol.*, S176, 1-87 (1963).
  - 13 F.M. Wiener, R.R. Pfeifer, and A.S.N. Backus, *Acta Otolaryngol.*, 61, 255-269 (1965).
  - 14 D.A. Baylor, G. Matthews, and K.-W. Yau, *J. Physiol. (Lond.)*, 309, 591-621 (1980).
  - 15 D.A. Baylor, T.D. Lamb, and K.-W. Yau, *J. Physiol. (Lond.)*, 288, 613-634 (1979).
  - 16 D.A. Baylor, T.D. Lamb, and K.-W. Yau, *J. Physiol. (Lond.)*, 288, 589-611 (1979).
  - 17 E.R. Lewis, E.L. Leverenz, and W.S. Bialek, *The vertebrate inner ear* (C.R.C. Press).  
in press
  - 18 G. Zweig, *Cold Spring Harbor Symp. Quant. Biol.*, 40, 619-633 (1975).
  - 19 E. de Boer, *Phys. Repts. (Phys. Lett. C)*, 62, 87-174 (1980).
  - 20 J. Lighthill, *J. Fluid. Mech.*, 106, 149-213 (1981).
  - 21 E. de Boer, *Hearing Res.*, 1, 253-281 (1979).
  - 22 W.M. Siebert, *J. Acoust. Soc. Am.*, 56, 594-600 (1974).
  - 23 G. Zweig, R. Lipes, and J.R. Pierce, *J. Acoust. Soc. Am.*, 59, 975-982 (1976).
  - 24 For this identification to be rigorous, we must have  $\langle VF \rangle = 0$ , which is not true. An exact correspondence can be made between Eq. (III.A.3.45) and the Hamiltonian of Eq. (D.2) after rotating about the *y*-axis through an angle  $\theta = \tan^{-1}(\langle VF \rangle / \epsilon')$ . This rotation is only important when  $\epsilon' \approx 0$ , and will be neglected except in this case.
  - 25 A.H. Zewail and C.B. Harris, *Phys. Rev. B*, 11, 952-963 (1975).
  - 26 The singularities of this equation when  $\epsilon' = 0$  are unimportant because we have explicitly discarded this case; cf. Ref. 24.
  - 27 W. Feller, *An introduction to probability theory and its applications* (Wiley, New York, 1968), Vol. 1.

- 28 R. Englman and J. Jortner, *Mol. Phys.*, *18*, 145-164 (1970).
- 29 W. Siebrand, *J. Chem. Phys.*, *44*, 4055-4057 (1966).
- 30 R.A. Marcus, *Ann. Rev. Phys. Chem.*, *15*, 155-196 (1964).
- 31 V.G. Levich, *Adv. Electrochem. Electrochem. Eng.*, *4*, 249-371 (1966).
- 32 J.J. Hopfield, *Proc. Nat. Acad. Sci. USA*, *71*, 3640-3644 (1974).
- 33 R. Englman, *Non-radiative decay of ions and molecules in solids* (North Holland, Amsterdam, 1979).
- 34 A. Sarai and T. Kakitani, *Chem. Phys. Lett.*, *77*, 427-432 (1981).
- 35 I.I. Abram and R. Silbey, *J. Chem. Phys.*, *63*, 2317-2328 (1975).
- 36 R.H. Austin, K.W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus, *Biochemistry*, *14*, 5355-5373 (1975).
- 37 J.L. Skinner, *J. Chem. Phys.*, *77*, 3398-3405 (1982).
- 38 R.F. Fox, *Phys. Repts.*, *48*, 179-283 (1978).

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT  
LAWRENCE BERKELEY LABORATORY  
UNIVERSITY OF CALIFORNIA  
BERKELEY, CALIFORNIA 94720