TRANSIENT TRANSPORT AND OPTICAL STUDIES
OF CHALCOGENIDE GLASSES

by

DONALD PAUL MONROE

B.S., Physics, Massachusetts Institute of Technology
(1980)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
January, 1985

© Massachusetts Institute of Technology 1984

Signature of Author

, Department of Physics

Certified by

Marc Kastner
Thesis Supervisor

Accepted by

George Koster
Chairman, Department Committee

JAN 25 1985
LIBRARIES
ARCHIVES
TRANSPORT AND OPTICAL STUDIES
OF CHALCOGENIDE GLASSES

by

DONALD PAUL MONROE

Submitted to the Department of Physics on
January 11, 1985 in partial fulfillment
of the requirements for the Degree of
Doctor of Philosophy in Physics

ABSTRACT

Amorphous Chalcogenide Semiconductors, primarily Arsenic Selenide, were studied using transient photocurrent and transient photoinduced absorption measurements. This work comprised three main subjects.

The transient photocurrent (TPC) measurement has been extended to substantially lower temperatures than in previous work; this extension revealed a transition to a $1/t$ time decay below 200K. This transition was explained as a change in the thermalization process for carriers in band tails, from thermal activation to hopping directly between localized states. A model for this process was developed, in which the competition between thermal energy and wavefunction overlap results in current being carried at a transport energy within the band tail. This model has many important implications for dc as well as transient transport.

The dependence of the TPC on excitation energy was studied for a broad range of sub-band-gap excitation energies. The TPC quantum efficiency is essentially constant throughout the exponential absorption edge, and drops off only slowly at lower energies. The significance of this result for models of the excitation process is discussed.

A detailed survey of transient photo-induced absorption was performed, covering a broad range of photon energies and temperatures. In addition, the importance of steady-state background excitation was demonstrated and explored using strong bias illumination. The results show that there are several separate contributions to the spectra, and their identification and significance are discussed. Among the observations is the first measurement of transient photo-induced bleaching of the "metastable" induced absorption.

Thesis Supervisor: Marc Kastner
Title: Professor of Physics
TABLE OF CONTENTS

ABSTRACT ............................................................................ 2

TABLE OF CONTENTS ............................................................... 3

CHAPTER 1: INTRODUCTION TO AMORPHOUS SEMICONDUCTORS
  1.1 Large-scale Electronic Structure........................................... 8
  1.2 Intermediate-Range Order ................................................. 11
  1.3 Sources of Disorder ......................................................... 15
  1.4 Effects of Disorder— the Anderson Model ......................... 18
  1.5 Extended States ............................................................... 20
  1.6 The Mobility Edge ........................................................... 23
  1.7 Localized States and Band Tails ........................................... 24
  1.8 Overview of Standard Model .............................................. 27

CHAPTER 2: EXPERIMENTAL APPARATUS
  2.1 Transient Measurements .................................................. 30
  2.2 Samples ......................................................................... 31
  2.3 Equipment in Common for All Experiments ....................... 32
  2.4 Photocurrent Measurements ............................................. 35
  2.5 Induced Absorption Measurements .................................... 36
  2.6 Photoluminescence measurements .................................... 37

CHAPTER 3: INTRODUCTION TO MULTIPLE TRAPPING
  3.1 The Multiple Trapping Problem ......................................... 39
  3.2 Time Dependent Occupation Number ................................... 41
  3.3 Trapping and Release Rates ............................................. 42
  3.4 Density of States (DOS) ................................................... 46
  3.5 Trapping-limited decays .................................................. 48
  3.6 Quasi-equilibrium Decays ............................................... 50
  3.7 Emission Limited Decay ................................................ 54
  3.8 High Temperature decay (T>T0) ......................................... 55
  3.9 Bimolecular Recombination and Saturation ....................... 60
  3.10 Repetitive Pulsing and Steady-State Background .............. 62

CHAPTER 4: OPEN QUESTIONS IN MULTIPLE TRAPPING
  4.1 Significance of Trap Parameters ....................................... 66
  4.2 Exponential DOS ........................................................... 72
  4.3 Effect of Electron-phonon Coupling on DOS ..................... 75
  4.4 Emission-Limited Current .............................................. 79
  4.5 Bimolecular Recombination ............................................. 81

CHAPTER 5: INTRODUCTION TO TRANSIENT PHOTO-INDUCED ABSORPTION
  5.1 Quantum Mechanics of Optical Absorption ....................... 84
  5.2 Transitions Between Extended States ................................ 86
  5.3 Transitions Involving Localized States ............................... 88
  5.4 Processes contributing to TPA .......................................... 90
  5.5 Room Temperature TPA: Thermalizing carriers ................. 92
  5.6 Shape of TPA Spectrum ................................................ 96

CHAPTER 6: LOW TEMPERATURE TRANSPORT ANOMALIES
  6.1 Low Temperature Photocurrent ....................................... 102
### APPENDICES

**APPENDIX A: HELPFUL HINTS FOR DYE LASER OPERATION**

A.1 General Discussion ........................................... 281
A.2 Individual Adjustments ...................................... 284
A.3 Diffraction Grating .......................................... 284
A.4 Oscillator Stage ............................................... 284
A.4 Preamplifier Stage ........................................... 291
A.5 Beam Expanding Telescope ................................... 292
A.6 Retro-reflections ............................................. 293
A.7 Amplifier Stage ............................................... 298
A.8 Power Tweaking ................................................ 300
A.9 Raman Shifter .................................................. 302

**APPENDIX B: CURRENT MEASUREMENTS**

B.1 Steady-state situation: Space Charge Relaxation .......... 305
B.2 Transient Measurements: Displacement Current .......... 308

**APPENDIX C: GEOMETRICAL ISSUES**

C.1 Photocurrent Geometry: Inhomogeneous Fields ............. 314
C.2 Photocurrent: Charging ...................................... 316
C.3 Induced Absorption- Transverse Geometry ................. 318
C.4 Induced Absorption- Longitudinal Geometry ............... 321

**APPENDIX D: CABLES ............................................. 324

**APPENDIX E: ACTIVE LOAD RESISTOR**

E.1 Basic Principles of Current Measurement .................. 329
E.2 Frequency Response of Active Load Resistor ............... 332
E.3 Physical Realizations of The PC Box ....................... 333
The average man takes everything
as either a blessing or a curse;
the warrior takes everything as a challenge.

-- Don Juan, as quoted by Carlos Castaneda

Whatever we do will be insignificant ...
but it is very important that we do it.

-- Mahatma Gandhi
CHAPTER 1
INTRODUCTION TO AMORPHOUS SEMICONDUCTORS

This thesis concerns various aspects of the behavior of electrons in amorphous semiconductors. To place this work in context, we will give a short introduction to the "standard" theoretical picture for amorphous semiconductors. This picture, in which the amorphous material is treated using the same approximations (one electron, nc electron-phonon coupling) as are used for crystals, has succeeded in explaining most observations. More exotic theories— involving polarons, correlation-induced gap, infrared divergences, and so forth— do not seem to be required. In addition to providing a context for the present work, we will mention some of the areas in which the standard picture remains unclear. The precise nature of the band conduction process, a major topic of this thesis, is one of those areas.

1.1 Gross Electronic Structure

The only order reliably known to exist in amorphous semiconductors is chemical: atoms tend to be found near other atoms in accordance with the rules of chemical bonding. For example, in As$_2$Se$_3$, most As atoms are bonded to three Se atoms, and Se atoms to two As atoms. Even this order is not perfect, of course, and deviations from this "ideal" bonding will constitute "defects" in the otherwise chemically perfect network. These defects and others, which give rise to electronic states different from the bulk of the material, will be discussed in Chapter 11. For the present we limit our attention to the "ideal glass".

The electronic structure of semiconductors is best discussed start-
Figure 1.1. Tight-binding picture of the chemical origins of the bands. When atoms are brought together, pairs of orbitals on adjoining atoms split into bonding/antibonding combinations, which are broadened into bands by other interactions. (a) tetrahedral semiconductors: basis can be envisioned as sp$^3$ hybrids.

In the tetrahedrally-bonded materials, represented by silicon, bonding and antibonding combinations of sp$^3$ hybrid orbitals can be regarded as the source of the valence and conduction band, respectively.
Figure 1.1(b) lone-pair semiconductors: basis is bonding, antibonding and lone-pair combinations of p-orbitals. Broadening into bands is not shown. Conduction band comes from antibonding levels, valence band from p-like lone-pair orbitals.

The splitting of the bonding and antibonding orbitals (~10eV) is reduced to ~1eV by longer-range interaction. Of course, the fact that these extra interactions are comparable to the original splitting is a warning that the simple tight-binding description should not be taken too seriously. This warning is reflected in the complex band structure of the tetrahedrals (for example, band extrema at low-symmetry points), indicating significant mixing of the bonding and antibonding orbitals. In addition, the technologically important a-Si:H is at best an alloy (at worst a grossly inhomogeneous material), and the role of hydrogen must
also be examined.

In the chalcogenides, represented by $\text{As}_2\text{Se}_3$, the situation is quite different, as first emphasized by Kastner (1972). Because there are more than four electrons on each atom, not all electrons will be involved in bonds. This reduces the advantage associated with hybridization (greater overlap) and the relevant orbitals are bonding/antibonding combinations of p-like orbitals on each atom, reflected in bond angles that are close to 90°. More importantly, because Se has six electrons, it will be only two-fold coordinated; four of the electrons will not participate in bonding at all. Two of these are in the s-shell and are much too deep to participate in the bonding, but two are in a p-orbital and have (at this level of approximation) the original energy of the p-shell.

These non-bonding, or "lone-pair" orbitals constitute the valence band of the chalcogenides, which are therefore called lone-pair semiconductors. Taking this observation at face value would seem to imply that the effective mass of holes would be extremely high, since the lone-pair orbitals on bonded neighbors are almost perpendicular and hardly interact; moreover, in $\alpha$-$\text{As}_2\text{Se}_3$ the neighbors of Se are As atoms and have no lone pairs. Such a conclusion exceeds the reliability of the chemical description, however, because interactions between more distant neighbors may suffice to make the effective mass comparable to the free-electron value. The high reactivity of the lone pairs does play an important role in stabilization of defects, as we discuss in Chapter 11.

1.2 Intermediate-Range Order

The basic features of an the amorphous semiconductor, on an energy scale of electron-volts (eV), are thus similar to that of a
corresponding crystal, since it both cases it is the chemical bonding which determines the electronic structure. Indeed, photoemission studies (Bishop and Shevchik, 1975) have substantially verified the predictions of the simple picture for the density of states (DOS) (although the lone-pair and bonding bands often overlap). This description is further supported by the observations of optical properties, which show transparency for photon energies below the band gap, and by the small, thermally activated electrical conductivity. These properties are similar for amorphous and crystalline semiconductors, and provide evidence that the large-scale band structure of both types of material is also similar.

Nonetheless, the simple chemical description does not suffice to tell us the properties of materials to the level of accuracy required to make it truly useful. The reason is that the electronic properties are sensitive to electronic structure on an energy scale of kT, which is hundredths of an eV, while the chemical energies are of order several eV. The electronic properties are also exponentially more sensitive to states close to the Fermi level, that is, within the bandgap. Since the chemical picture predicts no states in the gap, any states at all will be important.

The energy structure we have discussed so far has depended primarily on chemical ordering. The constancy of nearest-neighbor distances in amorphous semiconductors has been adequately demonstrated by scattering studies (Grigorivici, 1969). Variations in further-neighbor distances, however indicate a moderate spread (~10°) of bond angles, and the scattering experiments do not exhibit behavior corresponding to long-range correlations (microcrystallites). The scattering measurements
Figure 1.2. Structure of crystalline Arsenic Triselenide. The light spheres represent Se while the dark spheres represent As atoms. Note the \(-90^\circ\) bond angles the layered structure, and the helices of pitch four. (a) Cross section of layers, showing alternating helicity. (b) Top view of one layer; spirals run vertically.
are not very sensitive, however, and would not detect intermediate-range order on a length scale below ~20Å. The absence of any obvious order other than chemical has contributed to the widespread adoption of the continuous random network (CRN) for the structure, first proposed by Zachariasen (1932). In this model, bond lengths are preserved exactly, bond angles are allowed to fluctuate only slightly, but no other constraints are included. The conditions required for such an occurrence are complicated, however (Kastner, 1984), and several authors (Phillips, 1979; Sadoc and Mosseri, 1982; DiVincenzo, Mosseri, Brodsky and Sadoc, 1984, Brodsky and DiVincenzo 1983(a) and (b)) have proposed that significant IRO exists.

In particular, for a-As$_2$Se$_3$, there is some support for a locally layered structure, similar to that illustrated in Fig 1.2 for crystalline As$_2$Se$_3$. Experimental evidence for layering comes from X-ray scattering observations (Busse, 1984), of the "first peak" (longest wavelength) in the X-ray scattering. This peak, which is consistent with interlayer spacings in the crystal, exhibits anomalous temperature dependence near the glass transition. The theoretical work of Phillips is also interpreted as predicting an extended IRO. Zallen and Blossey (1976) and Zallen (1983) have enthusiastically endorsed the idea that As$_2$Se$_3$, and the related As$_2$S$_3$, exhibit quasi-two-dimensional behavior similar to the more ideal systems currently popular. In this regard, however, it is worth noting that although a random one-dimensional network can exist (as in a bowl of spaghetti or a linear polymer such as a-Se, a random two-dimensional network is impossible in 3-D, since any cross-linking between layers creates a topologically 3-D network. In any case, since the particular layered structure of the crystal results from
a complex competition between large energies favoring various configurations, it is hard to imagine how an extended layered structure could exist. Certainly the X-ray results show any IRO to be gone by a length scale of tens of Angstroms, so properties that average over longer length scales will not be affected by IRO.

1.3 Sources of Disorder

As shown in the preceding discussion, the basic features of the structure, important to a precise determination of the electronic structure, are very poorly understood. As a result, any discussion of the
electronic structure on a detailed level must be somewhat phenomenologi-
cal. Nonetheless, it is possible to identify a few of the salient con-
tributions, and to estimate their importance. For a-Si:H, this has been
done in a systematic way by Yoneyzawa and Cohen (1981); unfortunately I
know of no similar analysis for As2Se3.

As discussed earlier, the primary determinants of the electronic
properties are the nearest-neighbor distance (determining the
bonding/antibonding splitting) and the bond angle (which affects their
interaction). These parameters (and others) can be incorporated into a
simple hamiltonian:

$$H = \sum \varepsilon_i a_i^+ a_i + \sum_{<i,j>} t_{ij} a_i^+ a_j. \quad (1.1)$$

Here $\varepsilon_i$ represents the energy of the particular orbital (bonding, anti-
bonding, or lone-pair) and $t_{ij}$ represents the interaction between orbi-
tals, that is, the aspects of the full description that were neglected in
projecting onto tight-binding orbitals. If $\varepsilon_i = \varepsilon$ is a constant and
$t_{ij} = t$ for nearest neighbors and zero otherwise, this hamiltonian
yields a band of states extending from $\varepsilon-zt$ to $\varepsilon+zt$ where $z$ is the coor-
dination number. Thus the bandwidth $W=2zt$, a parameter we can compare
with experiment.

Even for these simple parameters, however, we are faced with a
source of disorder, which arises from the underlying topology of the
network. In its simplest form this topological disorder is viewed as
arising from the ring statistics, that is the distribution of
chemically-bonded rings with various numbers of members $m$. The reason is
that the two band extrema represent wavefunctions with alternate orbi-
tals either exactly in phase or exactly out of phase. For even-membered
rings, both are possible, but a wavefunction that changes sign an odd
number of times around a ring is not possible. This possibility affects the DOS at one band extremum. Interestingly, for chemically-ordered \( \text{As}_2\text{Se}_3 \), this source of disorder cannot contribute to the conduction band fluctuations, since each ring contains equal numbers of As and Se atoms and therefore an even number of bonds.

Other sources of disorder can and do contribute, however, which enter into the parameters of the Hamiltonian (1.1). These include bond length (affecting primarily the \( t_i \)'s), and bond angle and dihedral angle fluctuations (which affect primarily the \( t_{ij} \)'s). Since the relaxed structure reflects a balance between the various sources, it may be that the influence of these sources on the electronic properties are comparable (see Yonezawa and Cohen), at least for the valence band states, which are occupied during the relaxation. Furthermore, the diagonal and off-diagonal terms may well be correlated with one another (depending on their microscopic origins).

The parameters on nearby sites may also be correlated, resulting in longer-ranged fluctuations. This possibility has hardly been addressed in the literature, in part because of the difficulty of doing either calculations or simulations. Of course in the extreme long-length-scale limit, we simply have a percolation problem, in which classical carriers move in a fluctuating potential. The potential importance of such correlations can be seen by noting that local compression, say, will tend to spread over many atoms to reduce the total energy, depending on the growth conditions. Another possible source of long-range potential fluctuations is electrostatic fields, arising from local deviations from charge neutrality and extending over a screening length. Such fields may arise from local strain-induced polarization, or from the presence of
the electrically charged defects discussed in Chapter 11.

1.4 Effects of Disorder - the Anderson Model

In summary, the potential fluctuations through which the structural disorder influences the electronic states have many possible sources, differing in magnitude and length scale, and about which we know very little. Nonetheless, it is useful to examine the way in which disorder affects the electronic states, because disorder allows a new class of states, localized states, which are not found in perfect crystals. Of course, localized states in the form of impurity levels are found in real crystals; in this case too they are possible due to the lack of translational symmetry.

The localization of states in a disordered potential was first explored by Anderson in his classic paper (Anderson, 1958). The model he used was the simplest which exhibits localization: the hamiltonian of Eq. (1.1) is considered, with the sites forming a perfect lattice with a constant matrix element $t$ connecting nearest neighbors, 0 connecting more distant neighbors. The disorder appears only in the site energies $\epsilon_i$, which are chosen from a probability distribution of width $V$. The properties of this system are then critically dependent on the relative magnitude of the disorder energy $V$ and the transfer matrix element $t$.

For zero disorder, $V=0$, we have simply a standard tight-binding model with bandwidth $B=2zt$, and we can parameterize the disorder by $\delta = V/B$. For large disorder, $\delta \gg 1$, the coupling between sites is negligible, and the eigenfunctions are just the localized basis orbitals on various sites $i$ having the various site energies $\epsilon_i$; the bandwidth is then just $W$. Anderson's paper addressed the question of the value of $\delta$ above which all states in the band are localized, as illustrated in Fig. 1.3.
Figure 1.3. Overall behavior of bands as disorder is increased. (from Ziman, 1983) At low disorder the bandwidth is unchanged, and but the disorder induces band tails. With increasing disorder the bandwidth increases, until finally, at the Anderson Transition, all states become localized.

In amorphous semiconductors, we are concerned with the weak disorder regime, with $\delta$ of order 0.1. Most states in the band are little affected by the disorder. It causes scattering between the states but does not appreciably affect their energy, although the van Hove singularities in the DOS, which depend on the long-range order, are washed out. Most importantly, the sharp cutoff at the band edges is broadened into a smooth "band tail". The states in the band tail arise from
particularly strong potential fluctuations that occur at a few isolated places in the solid; the states are therefore localized to these places.

As the disorder increases, the localization applies to a greater and greater fraction of the states until finally Anderson localization occurs, as in Fig. 1.3. This is reasonable: in order for neighboring sites to mix together to form extended states the coupling (t) must exceed the typical energy spacing (V). Much of the literature on localization concerns this limit. For amorphous semiconductors it is more useful to discuss band-tail states in a framework which relies on the small magnitude of the disorder relative to the bandwidth.

1.5 Extended States

Because the disorder is small, only states near the band edge of the non-disordered material are involved in the band-tail states; these states can be described in the effective mass approximation. Fluctuations in $\varepsilon_i$ and $t_{ij}$ then correspond to fluctuations in the local potential $V(F)$ and the effective mass $m^*(F)$. We will restrict our attention to potential fluctuations. Schrödinger's Equation is then

$$- \frac{\hbar^2}{2m^*} + V(F) = E - E_0^C$$

(1.2)

Here $E_0^C$ is the band edge of the appropriate effective mass Hamiltonian, illustrated in Fig. 1.4. The appropriate average Hamiltonian depends on the average parameters of Eq. (1.1) as well as the disorder itself. However, the question of the appropriate averaging procedure, be it the Coherent Potential Approximation (CPA) or something else, is not relevant to our discussion here, as long as it is possible to cast the results in the effective mass form of Eq. (1.2).
Consider first states deep in the band, $E>V$, where $V$ characterizes the fluctuations $V(r)$. The solutions are superpositions of the plane waves on a thin shell having $\hbar^2 k^2 / 2m^* = E - E_c^0$, as shown in Fig. 1.5(a). The contribution of various plane waves within this shell is a random function depending on the particular form of the potential fluctuations. The envelope, however, is well defined and Lorentzian, corresponding to the lifetime broadening of the energy of the state (Thouless, 1975). The thickness of the shell is just $\Delta k = 2\pi / l$, where $l$ is the scattering length.

As the excess energy $E - E_c^0$ gets smaller, the wavevector $k$ gets smaller, and at a sufficiently low energy the thickness of the shell is comparable to its radius. At this point, the plane wave description is
Figure 1.5. Representation of energy eigenstates in $k$-space. (a) For weak scattering, $k\lambda \gg 1$, the eigenfunctions are distributed everywhere on the equal-energy shell, with a thickness $1/\lambda$. (b) As the wavelength becomes as short as the scattering length this description as a shell becomes inappropriate, and localization sets in.
no longer useful, and a new description, in terms of localized states, is more appropriate. The energy at which this occurs is given by the Ioffe-Regel criterion, kl-1 (Ioffe and Regel, 1960).

1.6 The Mobility Edge

The nature of wavefunctions near the mobility edge is dominated by the random potential fluctuations. In the early days it was common to assume that the wavefunctions had constant amplitude, but that their phase varied randomly from site to site. This assumption underlies one derivation of Mott's "minimum metallic conductivity," $\sigma_{\text{min}}$ (Mott, 1968). Mott concluded that the conductivity of an amorphous semiconductor should be given by $\sigma = \sigma_{\text{min}} \exp\left(-\frac{(E_c - E_F)}{kT}\right)$, where $E_c$ is the mobility edge energy, and $E_F$ is the Fermi energy. $\sigma_{\text{min}}$ is given by:

$$\sigma_{\text{min}} = C \frac{e^2}{na}$$  \hspace{1cm} (1.3)

where $a$ is a characteristic length for the wavefunctions, and $C$ is a constant, roughly .025. The modern theory (Abrahams, Anderson, Licciardello, and Thouless, 1979) suggests that there is no discontinuous change in the conductivity. This does not contradict Eq. (1.3), but requires that the length a diverge at the mobility edge, just as correlation lengths diverge at second-order phase transitions.

The analogy with second-order phase transitions suggests other features of the states near the mobility edge. For example, (Soukoulis and Economou, 1984) the wavefunctions vary strongly in amplitude as well as phase. The amplitude is large where the potential is deep, and eventually the states are localized in these regions.

The most potent analogy to localization is with percolation (see Thouless, 1983); indeed, localization can be regarded as quantum-
mechanical percolation. In localization in amorphous semiconductors, one takes the potential as fixed, and allows electrons to explore a greater volume of the solid as their energy increases. Precisely at the transition (the mobility edge), the wavefunctions are fractal in character, so the amplitude falls as a power of distance. For states further into the tail, this spatial variation is multiplied by an exponentially decreasing envelope, with a localization length that diverges as the edge is approached. Anderson (1972) calculated that this divergence should go as the 0.6 power of the energy difference. Unfortunately, the implications of these interesting phenomena for the current carried in states near the mobility edge is not easy to determine.

For potential fluctuations occurring on long length scales, the percolation analogy is even more appropriate: as long as the length scale is long compared to the length over electron wavelength, the motion of the electron can be treated classically. The localization problem then is the percolation problem, as stressed by Cohen (1970) and Cohen and Jortner (1973). Because of the diverging size of the wavefunction at the edge, all potentials will eventually look short-ranged, and localization will take over sufficiently close to the edge ("percolation crosses over to localization," in phase-transition language). This crossover may be irrelevant experimentally, however, depending on the energy range over which localization is important.

1.7 Localized States and Band Tails

The decreasing localization length of the wavefunction for states further from band center represents an increasing tendency toward localization for deeper states. The power-law behavior close to the mobility edge does not continue, however, because the localized states tend to
avoid regions of anomalously repulsive potential. This avoidance becomes important when the localization length decreases to the length scale over which the potential fluctuates.

Within the band tail a competition between kinetic and potential energy determines how well a wavefunction is able to take advantage of a given potential fluctuation. The form of the tail thus depends on the spatial behavior of the fluctuations, as well as their distribution of magnitudes. We will not explore this subject in detail, but Fig. 1.8 illustrates the basic features of the competition. The lowest energy state for a given potential fluctuation represents a balance between the average potential energy over the size of the wavefunction, $R_o$, and the kinetic energy, normally taken proportional to $R_o^{-2}$.

![Diagram](image)

**Figure 1.7.** Competition between kinetic and potential energy. The kinetic energy grows as the state is made more localized, but so does its ability to take advantage of a short-length-scale potential fluctuation.

The importance of kinetic energy depends on the length scale over which the potential fluctuates compared to the size of the wavefunction. Kane (1963), and much later Eggarter and Cohen (1970) examined the
situation for very long length-scale fluctuations. In this case the kinetic energy necessary to localize the wavefunction in any given potential well is negligible; the DOS is then a convolution of the local DOS \( \propto (E-V(R))^{1/2} \) with the probability distribution of potential fluctuations. This latter distribution is often assumed to be Gaussian, which is the appropriate form if many independent sources are contributing to the potential. The rapidly decreasing probability of finding a more negative potential fluctuation means that this probability dominates the convolution, and the DOS is thus Gaussian. If we define an exponent \( n \) and width \( w \), in

\[
g(E) = \exp \left[ -\left( \frac{E^0 - E}{w} \right)^n \right]
\]

the long-range Gaussian potential yields \( n=2, w=V \).

In contrast, Zittartz and Langer (1966) analyzed a white noise potential, which has fluctuations on all length scales, and found \( n=2-d/2 \), where \( d \) is the Euclidean dimension of space. In this case the kinetic energy is playing a major role. Halperin and Lax (1966, 1967) analyzed a more realistic potential and found an apparent exponent that varied continuously between 1/2 (for \( d=3 \)) and 2, depending on the size of the wavefunction compared to the potential length scale. In addition, the smaller size of deeper wavefunctions means that the crossover from large to small wavefunction (short-range to long-range potential) occurs as the energy moves deeper as well. The crossover is gradual, however, and experiments will tend to measure only one exponent \( n \).

This last result was very useful for theorists because more-or-less exponential tails \( (n=1) \) are ubiquitous. (We will discuss later in the thesis both the optical absorption and transient photocurrent evidence
for this statement.) In the Halperin and Lax picture, this is taken to be an accident depending on the particular magnitude of parameters involved. Recently, Soukoulis, Cohen, and Economou (1984) have presented a theory in which n=1 is not an accident, but an intrinsic feature of the tails, for a short-range potential. The basic idea is exactly the same as the Halperin and Lax theory, except that the kinetic energy is calculated differently. The short wavelength fluctuations are taken to result in fractal variations in the wavefunction within a long-wavelength potential fluctuation, so that the kinetic energy is no longer proportional to $R_o^2$, but to a different power of $R_o$ (it is easier to compress a fractal wavefunction); this result in tails that are precisely linear for $d \geq 2$. There is a serious question whether the fractal effects should be important for deep states, however. In any case, the long-range picture is not altered by these arguments, and a crossover could be observed here as well.

1.8 Overview of the Standard Model

The overall picture for amorphous semiconductors is summarized in Fig. 1.7. The band gap definitely exists, but the absence of sharp band edges makes a precise numerical value for the energy less meaningful.

The existence of band tails of localized states near the band edges has a major influence on the behavior of carriers thermally or optically excited across the gap. This is because the proximity of the tail states to the Fermi level leaves them exponentially more occupied than states deeper in the bands. The trapping of carriers into these states causes the unusual dynamics discussed in Chapters 3, 4 and 5.

At reasonably high temperatures, conduction occurs near the mobility edges $E_C$ and $E_V$, but the precise nature of the conduction is not
Figure 1.8 General picture of the density of states in amorphous semiconductors. Illustrating the bandgap and the band tails, as well as possible states deep in the gap.

clear. We will discuss this in great detail in Chapters 7, 8 and 9. We will see that the introduction of hopping directly between band-tail states modifies the picture of conduction in some important ways. This process also modifies the transient measurements, and the experiments that first showed the need for its inclusion are discussed in Chapter 6.

Almost all experiments described in this thesis depend on the optical excitation of carriers. Unfortunately, the nature of this excitation process in amorphous semiconductors is poorly understood. Some
interesting experimental results are discussed in Chapter 10.

Because of the exponential dependence of occupation on energy, states deep in the gap will have an important effect on the electrical properties. The apparently high density of gap states in the Chalcogenide Glasses is the major problem for their technological application as solar cells. Some experiments that give information about these important states are discussed in Chapters 11, 12, and 13. Unfortunately, the implications of these later experiments are not as clear as those in the earlier chapters; partially this reflects the many parameters involved. The final chapters will be primarily an overview of some of the interesting observations, and an survey of the parameter space.
CHAPTER 2

EXPERIMENTAL APPARATUS

The preceding chapter introduced the minimal description of amorphous semiconductors, comprising valence and conduction bands with band tails extending into the gap. The best evidence for such band tails comes from transient photocurrent measurements, as discussed by several authors (see for example Schmidlin (1977) or Noolandi (1977) for a-Se, Orenstein and Kastner (1981) for a-As$_2$Se$_3$, Tiedje and Rose (1981) for a-Si:H). Other aspects of the model are also probed by this and other measurements. This chapter gives an overview of the experiments we use; many more details are discussed in the appendices.

2.1 Transient Measurements

The explosive growth in the number of transient measurements performed in recent years has partially been stimulated by technological developments. Specifically, the availability of convenient pulsed lasers, as well as digital transient recorders and computer processing capability, has made the acquisition and interpretation of temporal data vastly easier. In addition, there is a growing realization that transient measurements represent a new spectroscopy, if by spectroscopy we mean a method of distinguishing different states from one another on the basis of some parameter. In transient spectroscopies, the parameter is the response time of the states, and for amorphous semiconductors this additional spectroscopic variable has distinguished previously undistinguishable states. In addition, since the response time depends sensitively on the energy of the states, the temporal spectroscopy is closely
related to an energy spectroscopy.

A transient measurements consists of a sudden change in the conditions of the sample at a time t=0, while recording some observable property of the sample. For the experiments described in this thesis the sudden change is induced by a pulse of light from a tunable laser system. Observables include the conductivity, optical absorption, and luminescence. These various experiments provide complementary information about the changing distribution of carriers among the states.

2.2 Samples

All the experiments described here were performed on glassy samples. Most were on As$_2$Se$_3$, obtained in five nines purity from British Drug House. The material was placed in a quartz ampoule, which was evacuated to a pressure below one micron and sealed with a hydrogen torch. The ampoule was then heated to roughly 900K (600°C), at which temperature the low viscosity of the melt assured homogeneity. Quenching from these high temperatures resulted in a sample with large frozen-in stresses, which fractured easily. The sample used was therefore cooled to just above the glass transition temperature at 750K (450°C) and then quenched in air (pulled out of the oven). No systematic study of the quenching procedure was performed, since quench rate is generally thought to have little effect on the material properties. It should be noted, however, that Andrienssens, Hammam, Michiel and Marshall (1983) and coworkers have reported significant effects of quench rate, belying the conventional wisdom that glassy (as opposed to evaporated) samples are all the same. After cooling, the ampoule was broken open and the sample was cut and polished into a 0.5×0.6×0.7 cm$^3$ parallelepiped. The brick-shaped geometry provided a convenient optical window for optical
measurements and simultaneously a well-defined electrode geometry for current measurements (see Appendices B and C).

The Selenium sample was prepared in a similar fashion from six nines purity elemental Selenium. It was quenched directly from 900K (600°C) in a procedure designed to remove residual oxygen impurities. In addition, because of the poorer glass-forming ability of Selenium, it was quenched in liquid Nitrogen rather than air. Significant differences were noted between different Selenium samples with different polishing procedures, which were ascribed to surface crystallization of the sample that had undergone high speed polishing on a motorized wheel.

The Arsenic Sulfide sample was one prepared by Kazuro Murayama, who took pains to compensate for the higher vapor pressure of sulphur to insure stoichiometry. Both the selenium and arsenic sulfide samples were plates, roughly 2mm thick, equipped with painted-on, coplanar, graphite electrodes as described by Orenstein (1981).

2.3 Equipment in Common for All Experiments

The excitation consisted of a train of roughly 10ns long pulses at a 10Hz repetition rate produced by a Quanta Ray YAG laser/dye laser/Raman shifter apparatus, discussed in more detail in Appendix 1. This source can provide photons with energies almost anywhere in the range 0.5eV < Ėω < 5eV. In particular, we can excite with energies greater than the bandgap (Ėω > 1.9eV for a-As2Se3), presumably creating separated electrons and holes, or in the exponential absorption edge (Ėω ~ 1.7eV) which at least at low temperatures creates primarily excitons, or with even lower energies (Ėω < 1.5eV) to explore excitation in the low-energy absorption tail. For induced absorption measurements (see below), the laser beam was spread out and passed through an aperture of
known area, to provide a relatively uniform excitation density. In addition, a calibrated fraction of the beam was directly monitored, using a pyroelectric detector, so knowing the photon energy and the area of the aperture allowed a precise measure of the number of photons per unit area hitting the sample in each pulse. This is an important quantity for interpreting the magnitudes of the signals produced.

The YAG laser system provided much of the improvement in signal quality of the present work over that of Orenstein, who used a Nitrogen-laser-pumped dye laser. The nitrogen laser works only in the visible, and even there provides substantially less energy per pulse, -30μJ versus -30mJ for the YAG laser. In addition to the obvious loss in signal size, this reduction means that the experimenter often cannot afford the loss of intensity required to properly aperture the beam, resulting in further uncertainties in geometry. In the infrared the nitrogen laser is totally useless.

The samples were held usually held in a Janis optical-access dewar, which allowed the temperature to be varied between 2K and 300K by using an internal heater and liquid Helium as a cryogen. For many experiments, however, temperatures below 77K are not necessary, in which case it is cheaper and easier to use liquid Nitrogen as a cryogen. Unfortunately, the higher temperatures, as well as the higher heat capacity of Nitrogen, severely tax the heater internal dewar heater; for these situations temperature control was achieved with a homemade system described in part by Monroe (1980). For measurements above room temperature, a small copper oven wrapped with heater tape was used, using a commercial temperature controller.

Each of the measurements generates a signal equivalent to a current
source. The optimal conversion of the current into a voltage for a given
time range is discussed in Appendix E. The temporal characteristics of
the laser restrict useful transient data to times longer than a few tens
of nanoseconds. Several amplifiers in our lab provide adequate time
response, including the PAR 115, LeCroy, and Comlinear; only the PAR 115
has a high impedance input. For times longer than 0.5µs, a more versa-
tile, higher input impedance amplifier such as the PAR 113 or the ITHACO
1201 provides highly variable gain as well as variable high- and low-
pass filtering for minimal noise.

Following amplification, the signals resulting from successive
excitation pulses were averaged using one of two systems. For short time
delays, a Bicron 6500 waveform digitizer, together with a Tracor 1505
signal provided averaging for data at repetition rates up to 20Hz. Only
two waveforms can be stored on this machine, so the data must immedi-
ately be transferred to the computer, which is a MINC System 23. For
times longer than 0.5µs there exists the option of using a Nicolet 4094
digital oscilloscope. Although this instrument provides averaging at
only 3Hz repetition rate, the sacrifice of signal to noise is sometimes
justified by its greater flexibility. For example, numerical information
can be obtained directly from the screen, yielding quicker feedback on
the results of some measurements, and data need not be transferred to
the computer immediately. In addition, it is possible to average, simul-
taneously with the signal, the output from the pyroelectric detector,
and thus obtain a very reliable measure for the incident pulse energy
even in the presence of significant drifts or pulse-to-pulse fluctua-
tions. This instrument was always used for taking point-by-point TPA
spectra.
Digital data acquisition presents some special problems, related to the finite resolution of the analog-to-digital converter. These problems, which are particularly acute for the Biometion because of its six bit resolution, are discussed in Appendix F.

2.4 Photocurrent Measurements

For the transient photocurrent (TPC) measurement, the sample is equipped with electrodes, generally a few millimeters apart, and a voltage, usually 2000V, is applied to one of them. Free carriers created by the light pulse move in response to this applied field, and this motion shows up as a current in the external circuit; the manner in which this occurs is discussed in Appendix B. The current decays in a characteristic way with time, as the carriers make transitions out of high mobility states into low mobility states. Thus the TPC monitors the change in occupancy of higher-mobility states, as we will discuss in detail in the following chapters.

The TPC for amorphous semiconductors generally exhibits a more-or-less power law time decay \( i(t) = t^{-\eta} \), rather than the exponential decay more typically observed in crystals. This means that there is no single time scale for the experiment: decay is occurring on all time scales, and the experimental apparatus must be capable of making measurements with optimal signal-to-noise ratio on all time scales. Orenstein (1981) used a variable load resistor to achieve this purpose; the data presented here were taken using an "active load resistor" to achieve even better results. Further improvements are still possible, as discussed in Appendix E on the active load resistor.
2.5 Induced Absorption Measurements

In the transient photo-induced absorption measurement (TPA), we monitor changes in the transmission of a beam of sub-bandgap light passing through the sample, called the probe beam. While the TPC measurement is sensitive only to charged, mobile carriers, the TPA is sensitive to any of several excitations including mobile carriers, trapped carriers, and excitons. Of course, this sensitivity sometimes poses difficulties in interpretation. By monitoring the transmission for several different probe photon energies, one obtains spectral information about the excited species, which may allow one to distinguish the different contributions.

The probe beam used here was derived from a 75W tungsten-halogen lamp. This white light was pre-dispersed using a Jarrell-Ashe prism monochromator to lower the average power. For some measurements the average intensity of the beam was reduced still further by insertion of neutral-density filters, or by a mechanical shutter which reduced blocked the light during the period between pulses when data was not being taken. Because of the spatially varying spectral content of the prism monochromator output, significant care was necessary in aligning the optics to avoid spurious effects.

The prism monochromator slits were imaged onto the sample by a toroidal mirror, and the transmitted light was collected and refocussed by another toroidal mirror onto the input slits of a McPherson .3m grating monochromator. The grating monochromator provided the ultimate resolution of the measurement (typically 20nm) and also rejected scattered laser light. This latter function was particularly important when using low excitation energies, since for these energies the sample did not
filter out the excitation light.

The grating monochromator output was focused onto a semiconductor detector using an ellipsoidal mirror. For high energy probes a Si or Ge detector was optimal, but since the signal-to-noise was usually limited by the low-energy part of the spectrum, a cooled InAs detector, operated in photocurrent mode, was generally used for the whole spectrum. The optimal acquisition of photodetector response is essentially the same as that for photocurrent, and is discussed in detail by Higashi (1982).

In the past, the induced absorption signal has been plotted as a fractional change in transmission $\Delta T/T$, which is typically negative and of order $10^{-4}$. Since this change in the incident intensity is small, this is equivalent to a total change in absorption cross-section per unit area. Thus by normalizing $-\Delta T/T$ to the number of absorbed photons per unit area, we obtain the average change in absorption cross section $\Delta \sigma$ per absorbed photon. This is the quantity plotted in the figures, and it provides a more useful indication of the magnitude than $-\Delta T/T$, which is of course dependent on the incident intensity. Of course, it is difficult to be sure that illumination uniformity and geometry are perfect, so the absolute scales are uncertain to within factors of two. In addition, for very low energy excitation one must correct for the incomplete absorption of the excitation by the sample, introducing further uncertainties in the absolute scale. The relative magnitudes of data taken under the same conditions, and thus the shapes of the spectra and the time dependences, however, are much more reliable.

2.6 Photoluminescence measurements

Although most references to luminescence measurements in this thesis will be to other people's work, it is worth mentioning that the
TPA apparatus can also be used to measure luminescence, since it collects any light emerging from the sample and delivers onto a photodetector. However, the probe beam usually has a low divergence (large f-number), a system which collects all of the probe may still have a poor collection efficiency for luminescence (which is radiated into a large solid angle).

At low temperatures, the TPA collected will automatically include a contribution from the luminescence, which must be subtracted. Since measurement of the luminescence to be subtracted is done by blocking the probe beam, it is critical that the response of the detector be linear, so that the luminescence signal is independent of the total amount of light incident on the detector.
CHAPTER 3

INTRODUCTION TO MULTIPLE TRAPPING

It is now widely accepted that the transient photocurrent near room temperature arises from the phenomenon known as Multiple Trapping (MT) in a variety of amorphous semiconductors. (Silver and Cohen, 1977; Schmidlin, 1977; Noolandi, 1977) Our review of this process will therefore not follow the historical development, but will set out the framework that has been developed over the past few years. For a more detailed discussion the reader is referred to Orenstein, Kastner, and Vaninov, 1982, (hereafter OKV) or to Orenstein (1981). The approach taken here is that of Monroe and Kastner (1983), which can treat a wide variety of physical situations while retaining the physical intuition that comes with the simple argument given by Orenstein and Kastner (1981) and Tiedje and Rose (1981). The present chapter is largely theoretical, although examples from a-As$_2$Se$_3$ will be shown. The next chapter discusses the experiments in more detail, in particular some outstanding questions about the MT process.

3.1 The Multiple Trapping Problem

The Multiple Trapping (MT) problem addresses the dynamics of carriers in transport states interacting with a distribution of traps. By "transport states" we mean of set of high mobility states that are in internal equilibrium on the time scale of the measurement. For example, the extended states are normally among the transport states. In later chapters we shall see that when hopping between localized states is considered, the localized states close to the band edge are also transport
states. The transport states are separated from the deeper states, the true traps, by the transport energy, which we will take as the zero of energy. The energy of a trap is then negative, with a larger magnitude for a deeper state. We will limit our attention to carriers in a single band, for simplicity, the electrons in the conduction band; the results hold equally well for holes, with appropriate redefinition of parameters. Indeed, for $a$-$As_2Se_3$ time-of-flight measurements (Pfister and Scher, 1977) show that the holes are the only mobile carriers.

The transitions between each trap and the transport states are characterized by two coefficients: the release rate $\nu$ (in $s^{-1}$), and the trapping coefficient $b$ (in $cm^3/s$). These coefficients are defined by the equation for the occupation number $f$ of the trap:

$$\frac{df}{dt} = bn(t)(1-f) - \nu f, \quad (3.1)$$

where $n(t)$ is the number of carriers in transport states at time $t$. The essence of MT is the absence of transitions between traps, so that each trap interacts directly only with the transport states. It is this fundamental simplicity that makes rigorous analysis of the MT problem possible.

In equilibrium, $f = \exp[-(E-E_F)/kT]$, where $E_F$ is the Fermi level. $n = N_0\exp(E_F/kT)$, where $N_0$ is the effective density of transport states referred to the transport energy $E=0$. Since in equilibrium the total rate of change is zero, it must be that

$$\nu = bN_0\exp[-|E|/kT]. \quad (3.2)$$

Thus the ratio between trapping and release coefficients is determined solely by the energy of the trap. The principle of detailed balance requires the ratio to have this value even if hopping between traps
is also important. The overall magnitudes of the trapping and release rates can be determined only through a detailed examination of the trapping process, however; their physical significance will be discussed in Section 4.1.

3.2 Time Dependent Occupation Number

The simplicity of the MT model is that the occupation of all traps at any time can be directly determined exactly from a knowledge of the number of carriers in transport states, \( n(t) \), at all previous times. For the present we will assume no saturation of the traps, that is, \( f << 1 \) for all traps. Saturation will be discussed in Section 3.9. Taking \( 1-f = 1 \), Eq. (3.1) becomes linear and can be integrated:

\[
f(t) = b \int_{-\infty}^{t} n(t') e^{-\nu(t-t')} dt'
\]  

(3.3)

This expression is easily understood: carriers fall into the trap at a rate \( bn(t') \), and then leak out with a constant release rate \( \nu \). Note that the form of this integral depends only on \( \nu \) and on the time dependence of the current; \( b \) enters only as a factor. Every trap follows an equation of the form (3.3), but the parameters \( \nu \) and \( b \) may be different for different traps.

We will describe the results of a transient photocurrent measurement, for which the generation of carriers in the transport states by a light flash can be considered to be a delta function at time \( t=0 \). The free carrier density \( n(t') \) contributes significantly to the integral in Eq. (3.3) only for a time \( \nu^{-1} \) preceding \( t \). Eq. (3.3) then exhibits two interesting limits, depending on the release time \( \nu^{-1} \), compared to the time \( t \) since the carriers were introduced.

**Fast states:** For states with a rapid release rate, \( \nu t >> 1 \), \( n(t') \)
will be essentially constant during the entire interval during which the occupancy was determined. The occupancy is therefore

$$f(t) = \frac{n(t)}{N_c} e^{-E/kT}$$  \hspace{1cm} (3.4)$$

This has a Boltzmann form, because these shallow, rapidly releasing states have come into equilibrium with the \textit{instantaneous} value of \(n(t)\).

\textbf{Slow States:} When \(vt \ll 1\), the exponential expression in Eq. (3.3) is always equal to one. Thus

$$f(t) = b \int_{-\infty}^{t} n(t') dt'$$  \hspace{1cm} (3.5)$$

These slowly releasing states have only experienced capture events; any carriers that have been captured by them are still trapped.

The full expression for the occupation is plotted in Fig. 3.1. The two limits represented by Eqs. (3.4) and (3.5) are given by the two straight lines. Since the shallow state occupancy is determined by the present \(n(t)\) while the deep state occupation is the integral of \(n(t')\) over all previous times, however, the relative magnitude of these two parts depends on the actual time dependence.

\textbf{3.3 Trapping and Release Rates}

Orenstein and Kastner (1981) and Tiedje and Rose (1981) used the form of the occupation number to deduce the TPC for the special case of an exponential DOS. For a more general discussion it is useful to examine the \textit{trapping and release} from traps, as well as their occupation, in the spirit of Monroe and Kastner (1983). The exact trapping/release rate \(df/dt\) can be determined from Eq. (3.3) for any \(n(t)\) and \(v\). The resulting expression can be written in the form

$$\frac{1}{b} \frac{df}{dt} = e^{-vt} n(t) - v \int_{0}^{t} [n(t') - n(t)] e^{-v(t-t')} dt'$$  \hspace{1cm} (3.6)$$
These terms correspond to trapping and emission, respectively, and will be discussed in turn. They are plotted in Fig. 3.2.

The trapping term is straightforward: carriers are trapped at a rate \( b_n(t) \). The trapping into fast states is not important because these states are in equilibrium with the transport states so the emission and trapping effectively cancel; this is the reason for the exponential cut-off for fast states. The somewhat arbitrary division of \( df/dt \) into
Figure 3.2. Emission and trapping rates, as separated in Eq. (3.6). Trapping is important only for states with rates slower than $1/t$. Emission is most important for states which are just becoming fast enough to emit; it depends on the history of the decay.

Trapping and emission terms were chosen specifically to exhibit only the unbalanced part of the trapping and emission.

The emission term is more complicated, and exhibits two distinct limiting forms depending on the release rate. For $v \ll 1/t$, any excess carriers will still be captured, and the emission is simply proportional
to the release rate. As the rate becomes comparable to 1/t, the states have a chance to rid themselves of excess carriers, coming closer to equilibrium with the present n(t):

$$\frac{df}{dt}|_{\text{uncovering}} = v e^{-vt} \int_0^t [n(t') - n(t)] dt'$$  \hspace{1cm} (3.7)

This expression is in fact a lower limit on the emission rate, since $e^{vt'}$ is always greater than one in Eq. (3.6). Fig. 3.2 shows that this emission is sharply peaked at $vt = 1$, falling off exponentially for $vt > 1$. We refer to this as the "uncovering" emission, because it represents the emission from states that are allowed to emit their excess carriers as the time scale becomes slow enough.

In contrast, states with a rapid release rate have already emitted most of their excess carriers and are almost in equilibrium with the transport states. Since the remaining excess represents the inability of the states to follow changes in n(t) instantaneously, we refer to this part as the "lagging" emission. For $vt >> 1$, this emission has the form $\frac{df}{dt} = b \frac{dn}{\nu dt}$. The extent to which the occupancy lags behind is greater the slower the release rate, and is just sufficient to keep these shallow states almost in equilibrium with the transport states. This term would have exactly the same form (but opposite sign) if n(t) were increasing with time.

The relative magnitude of the lagging and uncovering terms depends on the decay of n(t). Specifically, for rapid decays, the integral of n(t') becomes more important than the value of n(t) itself or its derivative, and the uncovering term therefore becomes more important for $vt$ comparable to one or smaller. This effect can be seen for the different curves in Fig. 3.2. For slower decays, the lagging emission connects smoothly to the peak in emission; the occupation number also
connects smoothly, justifying the quasi-equilibrium occupancy argument of Orenstein and Kastner and Tiedje and Rose. For decays faster than $t^{-1}$, however, the emission is best described by the uncovering term throughout the peak region, and the lagging part of the emission is much smaller. This last feature will be important below, when we discuss various types of decay.

### 3.4 Density of States

We have shown that, given the time dependence of $n(t)$, we can determine the occupancy as well as the trapping and emission rate for any trap. But the observed $n(t)$ is a consequence of the total of the trapping and emission from all the traps, and therefore depends on how many traps there are of each type. It is this sensitivity to the distribution of traps that gives transient photocurrent measurements their great power.

The occupation number and its rate of change depend principally on the release rate $v$. Because of the exponential dependence of the rate on energy, however, it is useful to express the results in terms of the energy of the trap, $E$. From equation (3.2),

$$v = \nu_0 \exp[-|E|/kT] \quad (3.8)$$

where the prefactor

$$\nu_0 = \frac{b}{N_c} \quad (3.9)$$

contains the physics. It is often assumed that this prefactor is independent of energy; it is thought of as an attempt-to-escape frequency and therefore of order $10^{12} s^{-1}$. In the presence of strong multiphonon emission effects the prefactor will be smaller, as discussed in Section 4.1. In the hopping picture discussed in later chapters it will
also be reduced, because of the small overlap between localized wavefunctions. Nonetheless, the activated form, which is all that is really used in the following discussion, remains valid.

We have distinguished two limiting forms for the occupation number on the basis of release rate. By using Eq. (3.9), we can also make this distinction on the basis of energy. The crossover, at which $\nu(E) = 1/t$, occurs when the energy is equal to the demarcation energy $E_d$, given by

$$E_d = -kT \ln(\nu_0 t)$$  \hspace{1cm} (3.10)

As a result, we can refer to fast states as shallow and slow states as deep. $E_d$ moves to deeper energies logarithmically in time. The states near $E_d$ play a special role in the dynamics since they are changing from slow to fast, and we refer to them as the marginal states.

In order to determine the dynamics, we need to know the number of states at each energy, which is the density of states (DOS) $g(E)$. Then

$$\frac{dn}{dt} = \int_{-\infty}^{\infty} g(E) \frac{1}{b} \frac{df}{dt} \, dE$$  \hspace{1cm} (3.11)

where $(1/b) \, df/dt$ is a function only of $\nu(E)$, as given in Eq. (3.6). It should be noted that each state is reflected in the dynamics in proportion to its trapping coefficient $b$, as first shown by Pollak (1977). Of course, in the presence of saturation this result would no longer be correct, since saturation is sensitive to the absolute magnitude of $f$.

Several different classes of behavior can be observed, depending on the type of state (shallow, marginal, or deep) which are dominating the trapping and release. Since the trapping and release rates for each energy are determined from equation (3.6), the product of the DOS and these rates determines their importance. The various possible types of decay we will discuss are summarized in Table 3.1.
<table>
<thead>
<tr>
<th>Dominant emission</th>
<th>Dominant trapping</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Near $-E_d$</td>
</tr>
<tr>
<td>Shallow (lagging)</td>
<td>$n(t)=n(t_1)$</td>
</tr>
<tr>
<td></td>
<td>$\times \exp \left[ \frac{bN_e}{N_o} \int_{t_1}^{t} N_d(t') , dt' \right]$</td>
</tr>
<tr>
<td>Near $-E_d$</td>
<td>$n(t) \propto t^{-1+\beta \rho kT}$, $0 &lt; \beta \rho kT &lt; 1$</td>
</tr>
<tr>
<td>(uncovering)</td>
<td></td>
</tr>
<tr>
<td>Deep.</td>
<td>Not possible</td>
</tr>
</tbody>
</table>

Table 3.1. Types of Decay.

3.5 Trapping-limited decays

If there is a large number of states at shallow energies, free carriers will come primarily from these states, which are in thermal equilibrium with the free carriers. These states simply act as a buffer, increasing the effective number of states at the band edge from $N_c$ to

$$N_c = \int_{\infty}^{\infty} g(E) e^{-E/kT} \, dE.$$  \hspace{1cm} (3.12)

This definition is the same as that for $N_c$, except that for the latter quantity states with $E<0$ are not included. The shallow traps act to slow the rate at which carriers fall into deeper traps, since the carriers only spend a fraction of their time in the transport states. That fraction is the effective degeneracy factor $N_c/\bar{N}_c$.

If emission is truly dominated by the shallow states, then the lower limit of the integral is not critical, and $\bar{N}_c$ is not a function of time. The time dependence is thus determined by the number of deep
traps:

\[
\frac{dn}{dt}\bigg|_{\text{trapping}} = -bN_d(t)n(t) \quad (3.13)
\]

where, from Eq. (3.6),

\[
N_d(t) = n(t) \int_{-\infty}^{E_d(t)} bg(E) \exp[-\nu(E)t] \, dE
\]

\[
\approx n(t) \int_{-\infty}^{E_d} bg(E) \, dE.
\]

The last expression follows from the fact that the rate of trapping, \(bn(t)e^{-\nu t}\), is approximately \(bn(t)\) for deeper states, but is negligible for states much shallower than \(E_d\).

Since the shallow carriers come into equilibrium on a time scale faster than \(t\), \(n(t)\) will decay as fast as carriers are trapped into deep states:

\[
n(t) = n(t_s) \exp\left[-\frac{N_c}{N_c} \int_{t_s}^{t} bN_d(t')\,dt'\right]. \quad (3.15)
\]

Here \(t_s\) is some time at which the emission is dominated by shallow states.

The total number of deep states, \(N_d(t)\), decays with time, according to Eq. (3.14). However, if states much deeper than \(E_d\) are dominating that integral, the decay will be negligible and the trapping rate will be roughly independent of time. The decay will then be purely exponential, with a time constant \((bN_cN_d/\tilde{N}_c)^{-1}\). The shallow states increase the time constant because the carriers spend only a fraction of their time in the band.

In the deep trapping case the rate of removal of free carriers is independent of time. This description also applies to monomolecular recombination and to post-transit time observations in a time-of-flight
experiment, and the dynamics in these situations follows the same equations.

If the integral (Eq. (3.14)) which defines the number of deep states is dominated by marginal states, then the number of deep states will decrease with time. For marginal states near $E_d$ to dominate the integral, the density must be falling rapidly near $E_d$, and the number of deep states must decrease rapidly as $E_d$ gets deeper. The resulting current decay will thus be concave upward on a double-logarithmic plot. We will see an example of such a decay in Section 3.8.

3.6 Quasi-equilibrium Decays

In the regimes just discussed, the rate limiting step was trapping into deep states of carriers in shallow states. We now consider the case in which the number of such shallow carriers is limited by the rate at which they are emitted from deep or marginal states. The occupation of the shallow states then simply reflects a balance between emission and trapping.

The simplest case occurs when the emission is from deep states; this is only possible if there is a peak in the DOS deeper than $E_d$. Because the occupancy of, and thus the emission from, deep states does not change with time, the current is simply a constant, corresponding to a balance between the emission and trapping rates.

A more interesting situation occurs when the emission comes from marginal states. In this case, which has been treated by previous authors (Orenstein and Kastner, 1981; Tiedje and Rose, 1981; Orenstein, Kastner and Vaninov 1982). The variations of the current provide a spectroscopy of the states near $E_d$, since the current just depends on how many states are uncovered per unit time. For marginal states to emit
more carriers than shallow states, the DOS must not vary too rapidly (no faster than a Boltzmann factor).

The particular form of the decay depends on the variation of the DOS with energy, especially the variation at energies near the demarcation energy. It is convenient to introduce the logarithmic derivative of the DOS, which we define as $\beta_0$:

$$\beta_0(E) = \frac{\partial \ln g(E)}{\partial E}.$$  \hfill (3.16)

This quantity is positive for a DOS which decreases into the gap, as for example in a band tail. If the DOS is exponential, then $\beta_0 = 1/kT_0$, where $kT_0$ is the energy width of the tail, in the notation of Orenstein, Kastner and coworkers (Tiedje and coworkers use $kT_0$). The form of the current transient at a given time will depend on the value of $\beta_0$ (actually $\beta_0 kT$) at the demarcation energy corresponding to that time, as we will see below.

We first discuss the case in which trapping is also dominated by states marginal states. This requires a decreasing DOS, $\beta_0 > 0$, since otherwise deeper states will trap more carriers. This form of decay is difficult to treat for an arbitrary DOS, because the emission rate (3.3) does not depend in a simple way on the history. However, an exact long-time solution is possible for a purely exponential DOS,

$$g(E) = \frac{N_L}{kT_0} e^{-E/kT_0}$$  \hfill (3.17)

where it must be true that $0 < T/T_0 < 1$, if the system is to be in the spectroscopic regime.

The current is a power law, $n(t) \propto t^{-\eta}$, the exponent of which can be determined quite simply: The current, $n(t')$, at a time $t'$, differs from $n(t)$ only by a factor $(t'/t)^{-\eta}$, at all times $t$. Because of this
absence of an intrinsic time scale, the form of the occupation number (Eq. (3.3)), is simply a function of $\nu t$, except for an overall factor: the $f(E)$ curve does not change shape, but simply translates through the DOS as $E_d$ gets deeper. The overall magnitude of the occupation number also changes, according to the requirement that the total number of carriers remain constant. The magnitude is thus inversely proportional to the DOS at $E_d$, $g[-kT \ln(\nu t)] \propto (\nu t)^{-\alpha}$, where $\alpha = T/T_0$. Since the Boltzmann factor describing the occupation of shallow states and transport states differs by a factor proportional to $(\nu t)^{-1}$ from the occupation number at $E_d$, and since the total number of carriers is conserved, the power law exponent is given by $1-\alpha$. When this exponent is inserted into Eq. (3.3) and integrated to give the total number of carriers, we obtain the magnitude of the current as well:

$$n(t) = \frac{\sin(\alpha \pi)}{\alpha \pi} \frac{N_c}{N} \frac{N(\nu t)^{-1-\alpha}}{N(L)}.$$  

(3.18)

Here $N$ is the total number of carriers, and $\gamma(\alpha, 1)$ is the incomplete gamma function. The numerical factor in braces is plotted in Fig 3.3; it goes to zero as $\alpha$ for $\alpha$ approaching zero, and as $(1-\alpha)$ for $(1-\alpha)$ approaching zero, but is otherwise unexceptional.

For $a-\text{As}_2\text{Se}_3$ (and to a lesser extent for $a-\text{Si}:\text{H}$), the TPC is an excellent power law, corresponding to an exactly exponential DOS. Data for $a-\text{As}_2\text{Se}_3$ at room temperature is shown in Fig 3.4. The observed value of $\alpha \approx 0.5$ corresponds to an energy width of the exponential DOS of $50 \text{meV}$, or $T_0 \approx 600 \text{K}$. The precision of this power-law behavior is discussed in more detail in Section 4.2.

The argument above is rigorous only for such an exact exponential DOS and a power law that extends back to zero time. However, the
integral that determines the total number of carriers is sharply peaked around $E_d$. Thus a DOS that varies more-or-less exponentially near $E_d$ will still give a slope $1-\alpha$ with $\alpha = \beta_0 kT$, with $\beta_0$ evaluated at $E_d = -kT\ln(\sqrt{\nu_0}t)$.

The contribution of different parts of the DOS is determined by the product of the DOS and the emission and trapping terms of Eq. (3.6). These products are sharply peaked near $E_d$, falling off as $kT_o$ on the deeper side, and as $kT_o/(T_o-T)$ on the shallower side. This provides a natural limit to the resolution of the measurement: $\beta_0$, for example, should be considered to be the logarithmic derivative of such a weighted-average DOS. Nonetheless, except for $T$ close to $T_o$, this resolution is quite good, and only very closely spaced discrete levels could escape detection. Marshall (1983) has pointed out that it is possible to
Figure 3.4 Transient photocurrent data for $\alpha$-$\text{As}_2\text{Se}_3$ at room temperature. Note the logarithmic axes.

Construct very special distributions of discrete states that yield approximately power law decays. The observed current shows no evidence for such a discrete distribution, however; it seems much more natural to assume that the featureless currents correspond to a featureless DOS.

3.7 Emission Limited Decay

As the number of deep trapping states decreases with time, eventually removal of carriers from transport states will be dominated by some process other than trapping at $E_d$, for example recombination. At that point, the decay will become faster, since the trapped states are no
longer reemitted, and very soon the lagging part of the emission \( \propto n(t) \) will be negligible with respect to the uncovering term \( \propto \int n(t^-) dt^- \). Since the decay is faster than \( t^{-1} \), as we shall see, the integral in Eq. (3.7) is sensibly constant once the decay is well underway, so the emission rate is just proportional to the product of the DOS and to the release rate at \( E_d \), which is \( 1/t \). The integral can be evaluated exactly, and the current is

\[
n(t) \propto t^{1+\alpha}(v_0 t)^{-1-\alpha}.
\]  

(3.19)

These features are exhibited in Fig. 3.5, which shows TPC decays in Arsenic Selenide at long times, at which the carriers begin to undergo monomolecular recombination (MR) with thermally excited carriers. The roll-over from a shallow slope to a steep slope is clear.

It is worth noting that for this deep-trapping case, it is no longer necessary that \( \beta e kT \) be less than one, so that a steep power law is the correct asymptotic form for any exponential DOS, in the long-time regime. Thus the post-recombination or post-transit-time decay can in principle be used to obtain spectroscopic information even at high temperatures. In practice, the rapid decay makes this difficult.

3.8 High Temperature Decay \((T>T_0)\)

The results of TPC measurements for \( T<T_0 \) strongly imply an exponential DOS (Eq. 3.17), and these phenomena have been explored in great detail by OKV, and by Orenstein (1981). For \( T>T_0 \), however, the current has traditionally been described as "nondispersive" (constant). Monroe and Kastner (1983) pointed out that even for \( T>T_0 \) the exponential DOS will manifest itself even here.

We continue to assume an exponential DOS, but we now examine the
case $T > T_0$. When the decay begins, there may be a transient associated with the initial carrier distribution, but for $t$ greater than a few $v_0^{-1}$ a situation such as depicted in Fig. 3.1 will apply. Since the DOS falls more rapidly than the Boltzmann factor away from $E=0$, the emission will be dominated by shallow states. Inserting Eq. (3.17) into Eqs. (3.11), (3.12), and (3.15), we obtain:

$$n(t) = N \frac{N_C}{N} \exp \left( \frac{N_L}{N} \frac{\alpha r(\alpha)}{\alpha - 1} (v_0 t)^{1-\alpha} \right)$$

(3.20)

so that
\[ \ln n(t) = \ln \left[ \frac{N_C}{N} \right] + \frac{N_L}{N_C} \frac{a}(\alpha) \exp \left[ -(\alpha-1) \ln (v_o t) \right]. \] (3.21)

This expression describes an exponential on a double logarithmic plot, decaying to a value \( N_C/N \) times the total number of injected carriers, if there is time to reach this level. The time required for \( n(t) \) to become only a factor of \( e \) greater than its final value is

\[ t_e = v_o^{-1} \left[ \frac{N_L}{N_C} \frac{a}(\alpha) \right] \frac{1}{\alpha-1}. \] (3.22)

This time is very short (of the order of \( v_o^{-1} \)) unless \( T \) is quite close to \( T_o \), at which temperature the duration of this transient becomes infinite. This extra transient provides for a continuous change from power-law decay for \( T<T_o \) to the constant long-time current for \( T>T_o \). The transient of equation (3.20) is similar to those observed by Silver, Cohen, and Adler (1981) in numerical simulations of MT.

The constant current, if it is reached, continues as long as deep trapping is negligible, which would be forever if only the states in the exponential were involved. At some time, however, some other mechanism will begin removing carriers from transport states; in the experiments of Thio et al. discussed below, this mechanism was monomolecular recombination. At this point, the decay becomes simply exponential, and the current rapidly decreases.

With the rapid decay of the current, the emission due to lagging will become less and less important, while the emission due to uncovering of deep states will become more important (see the discussion surrounding Eq. (3.7)). At a time not too long after the exponential decay begins, therefore, the lagging emission will cease being the dominant emission, and emission from uncovered carriers near \( E_d \) will dominate. We
Figure 3.5. Theoretical prediction for $T > T_0$ transients (from Monroe and Kastner, 1983a). The steep decay soon after onset is associated with a transition from emission by shallow traps to emission by states near $E_d$.

will then have the emission-limited power-law decay (steeper than $t^{-2}$!) of Eq. (3.19). The amount of time necessary before emission from marginal states supplies enough carriers grows with the ratio $N_c/N_L$.

The predictions for the time dependence are summarized in Fig. 3.5. These predictions have since received experimental support from the work of Thio, Monroe, and Kastner (1984). Some representative high-temperature transients are shown in Fig. 3.6. The decay immediately
Figure 3.6. High temperature \((T>T_0)\) decays in \(\text{a-As}_2\text{Se}_3\).

Note that the decay immediately following the onset of recombination is faster than that at longer times. Data courtesy Tineke Thio.

Following the onset of recombination is significantly steeper than \(t^{-2}\), which is the prediction for emission-limited current; we associate this part of the decay with the predicted exponential. At later times the data is consistent with a \(t^{-2}\) decay. Using the ratio between short and long-time decays, Thio et al. were able to estimate \(N_c/N_L\) to be around
20 at these temperatures.

3.9 Bimolecular Recombination and Saturation

The recombination processes we have considered so far have been monomolecular. That is, the rate at which free carriers recombine, which is proportional to the number of recombination centers, is independent of the number of added carriers. This is true when the injected carriers recombine with a thermally or optically excited background. If the number of electron-hole pairs injected per pulse is sufficiently high, however, recombination with other carriers excited by the same pulse will become dominant. In this bimolecular recombination (BR) case the number of recombination centers will decrease with time.

By "recombination center" we mean not simply a state in the gap but a normally filled state that has been emptied by the excitation (filled with an excited hole). An electron falling into such a center will disappear along with this hole and the system will be closer to equilibrium. According to this terminology (which differs from that of Rose (1978)) the total number of recombination centers is exactly equal to the number of excited electrons, that is,

$$N_R(t) = \int_{E_F}^{E_d} g(E)f(E,t)dE$$  \hspace{1cm} (3.23)

where

$$ \int_{E_F}^{E_d} g(E)dE .$$

OKV assumed that only free carriers are available for recombination, with a coefficient $b_R$ that will in general be different from the trapping coefficient $b$. The overall equation for the dynamics is then

$$\frac{dn}{dt} = \frac{dn}{dt}_{\text{emission}} - b_N(t) - b_RN_R(t) ,$$  \hspace{1cm} (3.24)
which is the same as the equation for monomolecular recombination (MR) or deep trapping, except that the number of recombination centers varies with time.

At long times, essentially all of the emitted carriers recombine rather being trapped in the decreasing number of deep states. The rate at which carriers are uncovered is just 1/t times the number of deep carriers, aside from a numerical factor, but the number of deep carriers is also the number of recombination centers, so the asymptotic form for the current decay is just 1/t, with a magnitude that is independent of the initial carrier density. The transition into this 1/t decay is broad, however, and is made even broader by the variations of density associated with the exponential penetration of the exciting light.

Recombination begins to affect the current decay only when the rate of trapping into deep states becomes less than that for recombination, that is,

$$bN_d(t) = b_R N_R$$  \hspace{1cm} (3.25)

or, using the expressions for the number of deep states,

$$\tau_R = \nu_0^{-1} \left[ \frac{bN_L}{b_R N_R} \right]^{1/\alpha}$$  \hspace{1cm} (3.26)

The recombination time is the same for both MR and BR, because at the onset of recombination the number of recombination centers is still at its original value. For BR, however, the number of recombination centers decreases with time once recombination begins. Moreover, an additional complication arises: saturation of the traps. It is easy to see that BR and saturation are intimately related, because since NR is equal to the number of deep carriers, then Eq. 3.25 is satisfied when the occupation \( f = b_R/b \), and since these coefficients may well be
comparable the occupation may well approach one before recombination occurs.

One expects a variety of possible behaviors, depending on the ratio \(b_R/b\), as discussed by OKV. If \(b_R \gg b\), then recombination begins at the time \(\tau_R\), long before the traps have saturated. On the other hand, for \(b_R \ll b\) the traps will saturate at a time

\[
\tau_{\text{sat}} \approx \nu_0^{-1} \left( \frac{N_L}{N} \right)^{1/\alpha}
\]

(3.27)

where \(N\) is the total number of carriers introduced. The decay of the current will then cease until recombination begins. This will occur at a time shorter than \(\tau_R\), since the saturation results in more carriers at high energies, available for recombination.

3.10 Repetitive Pulses and Steady-State Background

In actual experiments, the response observed is not really the result of a single pulse, as assumed above, but is the average response to many pulses. The sample is exposed to a series of pulses, and these can build up a steady-state background. The measured response is the additional current above that background. In later chapters we shall see that this background can sometimes have important effects; it is therefore useful to examine the expected effects of repetitive pulses within the simple model.

As we have noted throughout this chapter, the occupation of states depends on how their characteristic release time \(\nu^{-1}\) compares with some characteristic experimental time scale. For determining the importance of repetitive pulsing that time scale is the repetition time \(t_{\text{rep}}\) between pulses. States with a slow release rate, \(\nu_{\text{rep}} \ll 1\), do not respond significantly to the modulation by the laser; rather, they
attain an occupation determined by the average illumination. Returning to the full form for the dynamics (including saturation),

$$\frac{df}{dt} = bn(t)(1-f) - \nu f \geq 0$$ \hspace{1cm} (3.28)

so that

$$f_{ss} = \frac{1}{1 + \nu/bn}$$ \hspace{1cm} (3.29)

$$= \frac{1}{1 + \exp((E - E_{qf})/kT)}$$

In contrast to the demarcation energy $E_d$, $E_{qf}$ is a true quasi-Fermi energy: the states below $E_{qf}$ have $f < 1$.

The condition $\nu t_{rep} = 1$ occurs at an energy $E_{rep} = kT \ln(\nu_0 t_{rep})$. States deeper than $E_{rep}$ have a quasi-equilibrium occupation; they have come into equilibrium with the average carrier density $\overline{n}$. The modulation of their occupation by the pulsed nature of the light is a small fractional effect. In contrast, states shallower than $E_{rep}$ are able to come into equilibrium with the instantaneous free-carrier density $n(t)$ before the next pulse, and the modulation is as large as their occupation. Nonetheless, for shallow states that do not become saturated, the average occupancy still has the form of Eq. (3.29), in its Boltzmann limiting form, since in the absence of saturation the equations are linear. In contrast, if states do become saturated, their average occupancy depends on the exact time structure of the excitation; these states must be treated using the complete saturation description mentioned in the previous section.

We have shown that the occupation of states deeper than $E_{rep}$ are determined primarily by the average density $\overline{n}$. In a transient measurement, however, we measure not the total occupation $f$ but the change in occupation $\Delta f$ from its steady state value (Eq. (3.29)) induced by the
pulsing. This is given by

$$\frac{\Delta f}{dt} = b \left( \bar{n} + \Delta n(t) \right) \left( 1 - \left[ f_{ss} + \Delta f(t) \right] \right) - v \left( f_{ss} + \Delta f(t) \right)$$

(3.30)

$$= b \left( 1 - f_{ss} \right) \Delta n(t) - \left( b \bar{n} + v \right) \Delta f - b \Delta n \Delta f.$$

If either $\Delta f/f << 1$ (which will always be true for states deeper than $E_{rep}$) or $\Delta n/\bar{n} << 1$, (which will be appropriate if the quasi-Fermi level is shallower than $E_{rep}$), we can neglect the last term compared to the others. The dynamics then takes on the same linear form as used in most of this chapter, except that the coefficients are modified to $b^* = b(1-f_{ss})$ and $v^* = b\bar{n}/f_{ss}$.

For $f_{ss} << 1$ the modified values are the same as the original $b$ and $v$, so even if $\Delta f << f_{ss}$ the changes in occupancy follow the description used in the rest of this chapter. For saturated states, below $E_{qf}$, however, the modified coefficients are quite different. The effective cross section (and thus the effective DOS) is lowered by the fact that filled states are not available to capture carriers. More interestingly, the response time is changed from $v$, the rate at which carriers are emitted, to $v^* = b\bar{n}$, the rate at which carriers would be captured in the absence of saturation. If $b$ is constant, $v^*$ is just the release rate from states at the quasi-Fermi energy. This description leads to the important conclusion that the release rate at the quasi-Fermi level is the longest time in the problem; the system will come to quasi-equilibrium exponentially on this time scale.

In summary, the primary effect of repetitive pulsing is to introduce a steady-state background occupation corresponding to the average illumination level. This background may affect the development of the transients through nonlinear effects such as recombination. This is particularly important at low temperatures where the build-up will be much
more substantial, and we will discuss it in great detail in Chapter 12. The direct effect on the transient response of states, however, is limited to states below the quasi-Fermi level, whose effective density is reduced, and whose response time is shortened to the release time at the quasi-Fermi level. States that saturate after a single pulse are not describable by the quasi-Fermi distribution, and must be considered separately.
CHAPTER 4
OPEN QUESTIONS IN MULTIPLE TRAPPING

The previous chapter gave an overview of various experimental regimes exhibited in transient photocurrent measurements. In this chapter we will discuss the significance of the experimental observations in more detail, within the MT framework. OKV discussed in detail what parameters can be determined from the experiments. Our goal will be to reexamine some of those conclusions in the light of intervening experimental and theoretical results. The overall validity of the MT picture will not be addressed in detail, since this is the subject of later chapters.

4.1 Significance of Trap Parameters

We will first discuss the physical significance of the microscopic parameters governing release and trapping, which involve some very complicated and interesting physics which is not well understood even in crystals. The difficulty arises because transitions between traps and transport states involve emission or absorption of many phonons, since the energies involved are several tenths of an electron volt. Calculating the multiphonon transition rates requires a detailed knowledge of the local phonon modes at the trap, as well as the actual wavefunctions of both the trapped and the mobile electron. Such transition rates may well have a strong temperature dependence, over and above the temperature dependence of their ratio implied by the principle of detailed balance (Eq. 3.2). Detailed predictions would require a degree of knowledge about the localized states that is rarely available, espe-
cially in amorphous materials. For a review of multiphonon transitions, see, for example, Stoneham (1981) or Mott and Davis (1979), Chapter 3.

In spite of our ignorance about the exact values for the rates, however, we can classify transitions into two broad classes, depending on the degree of coupling between the electronic and vibrational degrees of freedom. More specifically, when the occupancy of a trap changes, the nearby atoms will rearrange, lowering the total energy by an amount $E_0$. If the parameter $S = E_0 / \hbar \omega_o$ (where $\hbar \omega_o$ is the energy of the relevant phonons) is large, then the system is strongly coupled, while for small $S$ it is weakly coupled.

In the strong-coupling regime, the important electron-phonon coupling is concentrated in a few vibrational modes of the lattice. These coupling of these modes to the electron is traditionally characterized by a configurational coordinate, $Q$. Fig 4.1 shows the total energy (electron plus phonons) of the system as a function of $Q$; the minimum energy lattice configuration (value of $Q$) depends on whether the trap is occupied.

Although the ability to represent the system in this way is reasonably general (for exceptions see Stoneham), $Q$ does not normally represent a single normal mode of the lattice, but some linear combination of modes. Energy put into the configurational coordinate may quickly dissipate into the other degrees of freedom. The dynamics of $Q$ is thus that of a oscillator with dissipation; such systems are surprisingly subtle. The active research on such systems has not yet been incorporated into the discussion of the transition rate, however.

In general, a transition between the transport state and the trap requires an energy $E_b$, the barrier energy, and is slower by a Boltzmann
Figure 4.1. Configurational-coordinate diagram representing trapping into a strongly electron-phonon coupled trap. The upper parabola is the total energy of the system when the electron is in a transport state; the lower parabola the energy when the electron is trapped.

factor than the value in the absence of a barrier:

\[ b = \frac{v_0}{N_c} \exp(-E_b/kT) \]  

(4.1)

The reverse transition is slowed by the same factor:

\[ v = v_0 \exp(-E_r/kT), \]  

(4.2)

where the release energy \( E_r \) is just

\[ E_r = |E| + E_b. \]  

(4.3)
Note that $E_b$ may be different for different traps, whereas $\nu_o$ is roughly a vibrational frequency. The traditional assumption of energy-independent $b$ is equivalent to $E_b$ being zero, or at least is the same for all traps.

In the original formulation by Mott (1938), $E_b$ was taken to be the energy from the minimum of the curve to the crossing point; quantum-mechanical tunneling through the barrier lowers the energy slightly (see Struck and Fonger, 1975). In the absence of detailed knowledge about the centers this is not particularly relevant. The tunneling also causes finite transition rates even at zero temperature: As temperature is lowered, the trapping coefficient is no longer activated, but approaches a minimum rate corresponding to transitions induced by the zero-point motion of the phonons (tunneling).

This behavior was observed experimentally by Henry and Lang (1977) in crystalline III-V semiconductors. Using DLTS spectroscopy they were able to measure the trapping cross sections as a function for a range of temperatures, for several different traps, as shown in Fig. 4.2. Although the room-temperature cross-sections differed by orders of magnitude, these variations can be attributed almost entirely to a range of barriers to trapping, which are as large as 0.5eV. The infinite-temperature extrapolations (proportional to $\nu_o$) for a variety of traps differed by only a factor of ten. This observation provides support for the assumption of a constant value of $\nu_o$ in Eq. (4.1). This assumption is particularly plausible for states with a similar microscopic nature, such as band-tail states.

The existing data for amorphous semiconductors, however, show no evidence of strong electron-phonon coupling for the states involved in
Figure 4.2. Temperature dependence of capture cross section of various traps in crystalline GaAs and GaP (from Henry and Lang, 1977).

multiple trapping. These observations include the absence of large deviations from the naive MT model, discussed in this section and the next, and the transient photo-induced absorption measurements discussed in Chapter 5.

One certainly would not expect strong electron-phonon coupling for the shallow band-tail states, since they are large and large states do not couple strongly to phonons. Deeper band-tail states ought to be small, however, and the absence of electron-phonon coupling is somewhat
surprising. A conclusive resolution of this question will depend on an understanding of the TPA spectra as discussed later, but the results do suggest at most a weak coupling.

In the weak-coupling regime, \( S \gg 1 \) the transitions result from the interaction of the trap with a large number of degrees of freedom; the configurational-coordinate diagram is not appropriate for this description. Theoretically, the temperature dependence of the rate follows the form of Kiel (1964) (see Robertson and Friedman, 1976), or Mott and Davis (1979), p.85:

\[
\nu_o \propto (1 + \langle n \rangle)^p
\]  

where \( p = |E|/h\omega_o \) is the number of phonons involved in the transition, and \( \langle n \rangle = [1 - e^{-h\omega_o/kT}]^{-1} \) is the thermal occupancy of the relevant phonon modes. For \( kT \gg h\omega_o \), \( \langle n \rangle \approx kT/h\omega_o \), and the transition rate at each energy has a power-law dependence on temperature, \( \nu_o \propto T^p \), corresponding over a narrow temperature range to an activation energy of \( pkT \). This weak dependence is not very important for our purposes. The energy dependence of the prefactor may affect the interpretation of some of the parameters of the DOS, however, as we will discuss in Section 4.3.

To estimate the effects of electron-phonon coupling on the results of the previous chapter, we will use the notation of the strong-coupling regime. A model-independent formulation is obtained by defining the release energy as simply a measure of the release rate:

\[
E_r = kT \ln(\nu_o/\nu)
\]  

We denote the density of states in release energy by \( g(E_r) \) to distinguish it from the true DOS \( g(E) \). The two quantities are related by

\[
g(E_r) = g(E) \frac{\partial E}{\partial E_r}
\]
\[ = g(E) \left[ 1 + \frac{\partial E_b}{\partial E} \right]^{-1} \]

which is well defined as long as \( \frac{\partial E_b}{\partial E} < -1 \) everywhere. (If this condition is not satisfied, a gap will open in the effective DOS; see Anderson, 1972.) For a weakly energy-dependent barrier the two densities of states will be of the same order of magnitude, and the energy dependence of their ratio will be insignificant compared to the exponential variation of the DOS. More importantly, the quantity actually measured is not \( g(E) \), but the product \( g(E) b(E) \), an effective DOS reflecting the degree to which the states participate in the multiple trapping process. If the parameters describing the energy dependence of the trapping were known, then we could convert the results of the previous chapter to exact results simply by making the appropriate substitutions.

4.2 Exponential DOS

In this section we examine the accuracy with which the DOS can be said to be exponential in \( a - \text{As}_2\text{Se}_3 \). Fig. 3.4 showed TPC data extending over a wider range of energy than had been previously observed. We shall see that this reflects a DOS that is accurately exponential over a wide range of energy, and is inconsistent with other forms that have been proposed for the band-tail DOS.

Obviously the construction of Fig. 3.4 required piecing together many (10) different transients, each acquired with a system optimized for a given time range. Although this sometimes required multiplying the adjoining data sets by a factor, the factor was determined only by the criterion that the magnitude be the same in the region of overlap; no preconception of the overall form influenced the matching.
A further experimental detail is that the data for times shorter than $10^{-2}$s was taken with $\tilde{\omega}_x = 1.600$eV, while for longer times $\tilde{\omega}_x = 1.400$eV. As discussed in Chapter 10, there is little effect of $\tilde{\omega}_x$ on the time decay in this photon energy range. In any case, one would expect that any differences between the two photon energies would show up in the present data. There may be some change in the time dependence near $10^{-2}$s, but it is small, and seems to make the curve less rather than more straight.

The observed photocurrent transient is obviously an excellent power law, covering more than eight decades in time. A least-squares fit to the data yields an exponent of $-0.48 \pm 0.01$ (one-standard-deviation error bars), corresponding to a value for $T_o$ of $570 \pm 10$K. The curvature $\partial^2 \ln i / \partial (\ln t)^2$ is $-0.002 \pm 0.004$, that is, zero, within experimental error.

According to Eq. (3.18), the TPC is inversely proportional to the density of states, so that,

$$g(E_d) \propto \frac{1}{t(\tau(t))} \quad (4.7)$$

Since the slope of the power law changes so little, it is unnecessary to include the dependence of the prefactor of Eq. (3.18) on the local power-law slope. The quantity $(t(\tau(t)))^{-1}$ is plotted versus $kT\ln(v_o t)$ in Fig. 4.3. The DOS appears exponential over a range of $kT\ln(v_o t)$ from 0.3eV to 0.75eV. Since the exponential width $kT_o$ is roughly 50meV, the DOS is exponential over an astonishingly large range of almost $10kT_o$.

The observation of an accurately exponential DOS is quite striking, and conflicts with most models of the density of tail states. As discussed in Chapter 1, theoretical treatments of band tails generally predict a more-or-less exponential form, but often with a different
Figure 4.3. Relative DOS in a-As$_2$Se$_3$. The vertical axis is $[i(t)t]^{-1}$, where $i(t)$ is the current, while the horizontal axis is $kT\ln(v_0t)$ with $v_0=10^{12}s^{-1}$.

We now examine the question of how well a DOS of the form of Eq. (5) can describe the data for $n=1$. If the form $g_n$ is to explain the data, the
The local logarithmic derivative of the DOS must give the measured $1/kT_o$:

$$\frac{\partial \ln g_n(E)}{\partial E} \bigg|_{E=E_o} = \frac{n}{w} \left( \frac{E_o - E}{w} \right)^{n-1} = \frac{1}{kT_o} \quad (4.9)$$

Since the data is accurately linear, the question is how small a curvature is compatible with the various forms. The curvature of the log-log plot of current for any $g_n$ should be

$$\frac{\partial^2 \ln I}{\partial (\ln t)^2} = (kT)^2 \frac{\partial^2 \ln g_n}{\partial E^2} \bigg|_{E=E_d} \quad (4.10)$$

$$= \frac{1-n}{kT_o(E_o - E_d)}$$

The curvature expected depends on the difference between the reference energy $E_o$ for the DOS and the energy $E_o - kT \ln(v_o t)$ at which the DOS is being measured. By using the observed limit on the curvature of $-0.002 \pm 0.004$, the minimum value for $E_o - E_d$ of 0.5eV for $n=1/2$, 2.5eV for $n=3/2$, and 5.5eV for a Gaussian, $n=2$; only the value of $n=1/2$ seems even remotely plausible. Since $E_o$ ought to be something like the $E^2_o$ of Fig. 1.4, one expects $E_o$ to be less than $E_C$, if anything, so even $n=1/2$ is unlikely. The case for an exponential DOS is quite strong.

4.3 Effect of Electron-phonon Coupling on DOS

The previous section ignored the possible distorting effect of electron-phonon coupling, discussed in Section 4.1. Obviously if one has the freedom to select the DOS and the capture coefficients independently, one can match the observed data in infinitely many ways. The only way to check these effects experimentally is to compare data at different temperatures, but even then one must make some assumption about the temperature dependence of the capture coefficients.
We first examine the observations using the activated-barrier approximation (strong coupling, with constant prefactor $\nu_o$). To compare transients at different temperatures, we simply plot the current at a given time and temperature versus $kT\ln(\nu_o t)$. If the barrier to trapping $E_b$ is the same for all states, then the curves for different temperatures will be parallel; this condition is satisfied for power laws that have a slope of $-(1-T/T_o)$ at temperature $T$.

![Diagram of transient photocurrent data at T=223K in a-As$_2$Se$_3$.](image)

Figure 4.4. Transient photocurrent data at T=223K in a-As$_2$Se$_3$.

For transient photocurrent in a-As$_2$Se$_3$ this temperature dependence was confirmed by OKV over a range up to 450K. Figure 4.4 shows more recent data at 223K, which shows a slope of $-0.63\pm0.01$, corresponding to a $T_o$ of 600\pm20K, in fair agreement with the room temperature value of
570±10K. Recently, Khan, Kastner and Adler (1983) have observed the 
α=1/T dependence in time-of-flight measurements on evaporated a-As$_2$Se$_3$, 
in contrast to previous work (Pfister and Socher, 1978, and references 
therein). Furthermore, the magnitude of the parameter $v_0$ (related to b 
by Eq. 3.2), at around $10^{12}$s$^{-1}$ is as large as is physically reasonable.

The data on a-As$_2$Se$_3$ thus preclude a **strong** variation of the capture 
coefficients. Nonetheless, it is possible that there exists a weak 
variation, and this may have a significant effect on some of the microscopic 
parameters. We will therefore discuss our results within the context of the weak-coupling theory, following Kagawa and Matsumoto (1983). 
The energy dependence of the release-rate prefactor includes, in addition to the temperature dependence of Eq. (4.4), an exponential dependence on the energy because of the improbability of p phonons being in the right place at the right time to induce the transition.

$$v_0 = v_o^* \exp(-\gamma |E|/\hbar \omega_o) (1 + \langle n \rangle)^p$$

(4.11)

$$= v_o^* \exp(-E/kT_p)$$

where $kT_p = \hbar \omega_o / [\gamma - \ln(1 + \langle n \rangle)]$. $\gamma$ is a constant of order two. Ogawa and Matsumoto assume that $kT_p$ is positive, and ignore its temperature dependence.

We continue to assume an exponential DOS, and examine the effects of the modification of the prefactor, following the discussion of Section 4.1 and ignoring the small difference between $\bar{g}$ and $g$: The demarcation energy occurs where the rate $v(E)=v_o(E)e^{-E/kT_p=1/t}$, so that

$$E_d = \frac{kT_p}{(T+T_p)} \ln(v_o t)$$

(4.12)

The carriers are not as deep, at a given time, as the straightforward
multiple-trapping model would predict. This delays the predicted time for saturation, apparently avoiding the problems described by Schiff (1981). The product $g(E_d) v_o(E_d)/N_o$ is the effective DOS, so the current is proportional to $t^{-1+\alpha}$, with

$$\alpha = \frac{T (T_o + T_p)}{T_o (T + T_p)}$$

(4.13)

$T_p$ need only be somewhat larger than $T_o$ to make no overwhelming change in the temperature dependence of the exponent, as shown in Fig. 4.5, but the value of $T_o$ is altered somewhat.

![Diagram](attachment:diagram.png)

Figure 4.5. Weak electron-phonon coupling corrections to the temperature dependence of the power-law exponent, Eq. (4.13), for fixed $T_p$.

In the presence of the weak electron-phonon coupling, the time for the demarcation energy to reach the Fermi Level is longer. One expects
to see an exponential decay of the excess current once the Fermi level is reached (see Section 3.10), so the absence of such an observation would otherwise be a significant puzzle. The situation is most clear in doped a-Si:H, for which the Fermi level is quite close to the bands, as inferred from the dc conductivity, but only the power-law decay due to recombination is observed. In general, the possibility that electron-phonon coupling alters the energy scale should always be considered when interpreting the importance of experimental data. These questions have not yet been fully explored.

The data already discussed allow us to rule out a large effect, however: the similarity of the apparent $T_o$'s at different temperatures means that $T_p$ is probably no smaller than $T_o$. When the data of OKV are included, the case becomes even stronger. In any case, although an exponential dependence of the rate on energy changes the energy calibration, it does not change the need for an exponential DOS if one is to explain power-law time decays. Although one could in principle invoke more complicated dependencies of the DOS and prefactor to explain the data, their conspiracy to yield an exact power law is quite implausible.

4.4 Emission-Limited Current

According to theory, the power-law exponents before and after the transit time or monomolecular recombination time should add to two, and this is clearly a reasonable description (see Fig. 3.5). However, an accurate determination of both of these slopes from a single transit requires data with good signal-to-noise covering at least five orders of magnitude in time, a very demanding requirement. It is therefore not surprising that there are many reports (Pfister and Soher 1978; Kristensen and Hvam, 1984) of slopes that add to less than two. The difficulty
in determining the slopes is particularly great in time-of-flight measurements, in which peculiar time decays (sometimes non-monotonic!) are observed near the transit time. These peculiarities are associated with space-charge regions near the contacts, and can be avoided through careful technique (Gibson and Kastner 1984; Khan, Kastner and Adler 1983). There may still be some discrepancy, however; for example, the slopes sometimes add to more than two.

This discrepancy may be explained by a model in which long-range potential fluctuations are responsible for the dispersive transport (Overhof, 1983, 1984). For times less than the transit time the behavior is similar to that for MT, but after the transit time the slope is steeper. The principal problem with this model, according to Overhof, is that it predicts a transit time which is inversely proportional to the field. This self-criticism cannot be correct; any mechanism that gives the correct time a tendency must give the correct field dependence, since the latter arises from the former. The failure to observe this may be an artifact of the Monte Carlo simulations, which does not inspire confidence in the other predictions.

Nonetheless, it is worth remembering that although the MT model provides an excellent explanation of the observations so far, other models which differ in some details may be more correct. Probably any model that involves a more-or-less exponential distribution of energies for carriers will give qualitatively similar results, whatever the details of the thermalization mechanism within this DOS. The long-ranged fluctuation model is one example; the hopping model to be discussed in later chapters is another. The extraordinarily pure power-law time dependence observed in a-As$_2$Se$_3$ places significant restrictions on the
4.5 Bimolecular Recombination

Fig. 4.6 shows data for bimolecular recombination in a-As₂Se₃, along with the theoretical prediction for \( b_R = b \) from OKV. The data match the theoretical curve rather well, especially considering that the horizontal adjustment of the data for different intensities is consistent with Eq. (3.26). Because of the absence of apparent flattening of the curve, OKV felt that there was no evidence for saturation, suggesting that \( b_R \) is at least comparable to \( b \). This conclusion is plausible, since a recombination center is more likely to be charged and to have a large, coulombic capture coefficient than a trap.

Closer examination of the data, however, reveals that each of the individual transients is straighter (changes slope less rapidly) than the theoretical curve. In other words, although the transition to \( 1/t \) decay is predicted to be quite broad (and is made broader by the exponential penetration of the light, which is included in the calculated curve), the actual transition is even broader. The agreement of the general features of the decay, especially the strong intensity dependence, suggests that the phenomenon is truly BR of carriers in an exponential DOS, but the details of the decay show some slight differences.

One possible explanation is that there is a distribution of recombination coefficients \( b_R \). To explain the breadth of the transition, this distribution would have to extend over more than an order of magnitude, and this is difficult (but not impossible) to explain in terms of trapping from plane-wave-like states. However, the recombination kinetics of carriers moving diffusively is surprisingly complicated, and exhibits
Figure 4.6. Bimolecular Recombination in a-As$_2$Se$_3$. The data for different intensities have been adjusted to provide the best fit to the theoretical curve with $b_R=b$. (From OKV).

long-time tails which may be affecting the time decay.

Alternatively, the possibility that the transport occurs via hopping, discussed in later chapters, suggests that there may well be a variety of tunneling distances (which enter into $b_R$ exponentially) between transport states and recombination centers. Since in BR every recombination center must be encountered by a carrier (in contrast to MR where there is an excess of recombination centers) either of these effects is a plausible explanation.
Another possibility is that there is some saturation of the traps, causing a slowing of the decay but not a completely constant current. This seems to be the case in a-Si:H, in which initial measurements (Hvam and Brodsky 1981), showed no saturation, but recent work by Kristensen and Hvam (1984) is interpreted as indicating saturation (as reflected in a superlinear intensity dependence). It is puzzling that there is still significant decay after the apparent saturation time. Such observations seem to invalidate the conclusions by Schiff (1981) of a major flaw in the MT model, which were based on the absence of trap saturation; the electron-phonon coupling discussed in the previous section may also be important here.

Hvam (private communication) has indicated that extreme homogeneity of illumination is required to observe these effects, and it is possible that similar care will reveal saturation in As$_2$Se$_3$. It should be easy to insure homogeneity using the YAG-laser-pumped dye laser. Since the magnitudes of the microscopic parameters determined from the BR measurements are used in interpreting several other results (they provide the only absolute calibration of the density, for example), this may be a worthwhile experiment.
CHAPTER 5

INTRODUCTION TO TRANSIENT PHOTO-INDUCED ABSORPTION

The transient photo-induced absorption (TPA) experiment has provided a wealth of information about the state in the gap of amorphous semiconductors. Because TPA is sensitive to a many different excitations, it has a wide range of applicability, but it is also difficult to interpret. In this chapter we introduce the experiment in the context of the thermalizing-carrier absorption associated with the MT process.

5.1 Quantum Mechanics of Optical Absorption

We begin with a short review of the quantum-mechanical description of optical absorption. Such a description is of course relevant to the primary excitation process, discussed in Chapter 10, as well as to the TPA experiment addressed here, and our discussion will at first be general enough to apply to both situations. In the remainder of the chapter we will specialize to absorption by states in the gap; a good review of the fundamental issues in that case is given by Dexter (1958).

In optical absorption, it is generally satisfactory to treat the radiation field classically, while treating the electrons quantum mechanically. The radiation then can be thought of as a perturbation of the "pure" electronic eigenstates of the system causing transitions through the perturbation Hamiltonian

$$H' = -\frac{e}{mc} \sum_i A(\vec{r}_i) \cdot \vec{p}_i$$

(5.1)

The sum extends over all electrons in the system, but it is usually adequate to envision the field as causing transitions between (renormalized) one-electron states, as we will do here.
The perturbation of Eq. (5.1) can be thought of as causing transitions between the unperturbed states of the system, which are described by (one-electron) states \(|j\rangle\) with wave functions \(\psi_j\) and unperturbed energies \(\varepsilon_j\). Fermi's Golden rule says that the rate of transitions from a particular occupied state \(|i\rangle\) to states with an energy near \(\varepsilon_f = \varepsilon_i + \hbar \omega\) (where \(\hbar \omega\) is the energy of the light), is

\[
\frac{df_i}{dt} = \frac{2\pi}{\hbar} |\langle i|H'|f\rangle|^2 \rho(\varepsilon_f) \tag{5.2}
\]

\[
= \frac{2\pi e^2}{m^2 c^2} |\vec{A} \cdot \langle i|\vec{p}|f\rangle|^2 \rho(\varepsilon_f) \tag{5.3}
\]

\[
= \frac{2\pi e^2 \omega^2}{\hbar c m^2} |\vec{A} \cdot \langle i|\vec{F}|f\rangle|^2 \rho(\varepsilon_f) \tag{5.4}
\]

where \(\rho(\varepsilon_f)\) is the total density in energy of final states \(|f\rangle\). The last equality follows from the fact that \(\vec{p} = \vec{r}/m\), and is valid only in the dipole approximation.

For centers in a dielectric medium, the local field where the transition occurs may be different from the average field by a ratio \(F_{\text{eff}}/F_0\). In addition, if the index of refraction \(n\) differs from one, there will be fewer incident photons per unit time, so the cross section will be altered by additional factors (Dexter, 1958). The effective field ratio may be as small as one (for a large center with the same \(n\) as the medium), or may be comparable to \(n^2\) (for a center with dielectric properties very different from the medium). Obviously this leaves some uncertainty in the magnitude, roughly a factor of \(n\) in either direction.

In relating this rate to experiments, two alternative quantities are convenient. The first is the cross section, which is the transition rate of Eq. (5.4) normalized to incident photon density:

\[
q_1(\hbar \omega) = \left[ \left( \frac{F_{\text{eff}}}{F_0} \right)^2 \frac{1}{n^2} \right] \frac{4\pi^2 e^2 \omega}{3c} |\vec{F}_{if}|^2 \rho(\varepsilon_f) \tag{5.5}
\]
The other convenient quantity is the total absorption coefficient $\alpha$ for all initial states $i$:

$$\alpha(\hbar \omega) = \int d\varepsilon_i \ g(\varepsilon_i) \ f(\varepsilon_i) \ \sigma_i(\hbar \omega) \quad (5.6)$$

The uncertainty in magnitude because of the effective field ratio also applies to the oscillator strength sum rule, which states that the integrated absorption by electron must be a constant, given by

$$\int \sigma(\hbar \omega) d\hbar \omega = \left[ \frac{F_{\text{eff}}}{F_0} \right]^2 \frac{1 - \frac{1}{n}}{\frac{2 \pi e^2 \hbar}{mc}} \quad (5.7)$$

$$\approx \left[ \frac{F_{\text{eff}}}{F_0} \right]^2 \frac{1 - \frac{1}{n}}{\frac{2 \pi e^2 \hbar}{mc}} \times 1.2 \times 10^{-16} \text{cm}^2 \cdot \text{eV}.$$  

Physically, the oscillator strength sum rule is a reflection of the fact that an electron subjected to a delta-function electric field will absorb an amount of energy that does not depend on the potential it is moving in, since the electron will not have time to "see" the potential. Since the Fourier transform of a delta function involves all frequencies equally, the sum rule follows.

In writing these expressions, isotropy has been assumed, so that the square of the projection of $\mathbf{F}$ along any direction averages to $\frac{1}{4} |\mathbf{F}|^2$. More importantly, the matrix elements are (rms) averages over all states included in the final state density $\rho(\varepsilon_f)$. Whether this is significant depends on whether the states involved are localized or extended, as we now see.

5.2 Transitions Between Extended States

It is customary distinguish between three types of absorption process: extended-to-extended, localized-to-extended, and localized-to-localized. It is normal to assert that localized-to-localized
transitions typically have exponentially small matrix elements; we shall see that this is not strictly true, and that the density of states plays the dominant role in determining which transitions are important.

The extended wavefunctions deep in the band have a magnitude that is essentially constant in space (we will discuss the fractal wavefunctions near the mobility edge shortly). For properly normalization, we must divide by the square root of the volume Ω over which we integrate wave functions:

$$\psi_x(\mathbf{r}) = Ω^{-1/2} \chi_x(\mathbf{r}).$$  \hspace{1cm} (5.8)

Throughout the following, the various $\chi(\mathbf{r})$ will be functions that have a maximum magnitude of order unity.

We first consider extended-to-extended transitions. The matrix element reflect the normalization of the wavefunctions:

$$r_{if} = \frac{1}{Ω} \int_Ω \chi_i^*(\mathbf{r}) \cdot \chi_f(\mathbf{r}) \, d^3r.$$  \hspace{1cm} (5.9)

If we consider starting with the integration volume Ω small and gradually enlarging it, then the matrix element will be constant as long as the initial and final wave functions maintain phase coherence, up to some length $\xi_{if}$. For $Ω > \xi_{if}^3$, various elements of volume $\xi_{if}^3$ will contribute randomly. The magnitude of $r_{if}$ will then grow only as $(Ω/\xi_{if}^3)^{1/2}$ times the contribution from the volume $\chi_{if}^3$; the latter falls as $Ω^{-1}$ due to the normalization. The square of the matrix element then falls as $1/Ω$. Since the number of final states $ρ(ε_f)$ is just $g(ε_f)Ω$, however, the absorption coefficient is independent of $Ω$, as it must be.

The absorption coefficient thus depends on the value the matrix element attains in the volume $\xi_{if}^3$. In crystals, $\xi_{if}$ is generally long enough that its value does not affect the matrix element; this will hap-
pen if $\xi_{1f}$ is much greater than the relevant wavelengths of the electrons. For an amorphous material, the large elastic scattering may cause $\xi_{1f}$ to be quite short (it will generally be shorter than the mean free path for either the electron or hole). Under these circumstances the normal $k$-conservation rule for optical transitions may be relaxed. It seems, however, that the conditions for this relaxation are intimately related to those for localization ($k\ell - 1$) discussed in Chapter 1; the apparent absence of the selection rule deep into the band (embodied in the Tauc rule, see Chapter 10) is surprising.

5.3 Transitions Involving Localized States

Localized states can be represented by

$$\psi_{\lambda}(\overline{r}) = \gamma_{\lambda}^{-3/2} \chi_{\lambda}(\overline{r}),$$

(5.10)

where $\chi_{\lambda}$ is again a function with maximum value one, and has an envelope that falls off like $\exp(-\gamma_{\lambda}|\overline{r} - \overline{R}_0|)$.

The cross section for localized-to-extended transitions is not affected by the localized nature of the initial state: the limited volume over which the integrand of Eq. (5.6) is significant is compensated by the normalization of the wavefunction. This result of course applies equally well to extended-to-localized transitions, by electron-hole symmetry.

Although localization does not affect the rate of excitation into extended states, it is commonly asserted that localized-to-localized transitions are suppressed because of small matrix elements. This appears to be false, as we now show. The matrix element between two localized states separated by $R$ will be exponentially small unless the centers are within $\gamma^{-1}$, the longer of the two wavefunction decay
lengths, of one another. For centers that are closer than this distance the larger wavefunction can be considered to be uniform over the volume of the smaller wavefunction. The matrix element will be that for localized-to-extended transitions, except that the normalization of the larger state is $\gamma_{<}^{3/2}$ instead of $H^{-1/2}$. The squared matrix element is thus proportional to $\gamma_{<}^{3}$. Since the probability of the two states being close enough (closer than $\gamma_{<}^{-1}$) to have a significant matrix element is proportional to $\gamma_{<}^{-3}$, these factors exactly cancel.

Thus although most of the matrix elements are indeed exponentially small, the average squared matrix element appears to unchanged by the fact that both states are localized. This argument should be correct if the states are the same size as well. The only reason localized-to-localized transitions are unlikely is because there is a smaller density of localized states; this is of course intimately connected to the localization.

One important issue which we have not addressed is the possibility of correlation between the initial and final state wave functions. For example, band-tail states will be localized in regions of large potential fluctuations. Extended states near the mobility edge undergo amplitude fluctuations (which gives them their fractal character) that will make them larger in regions of large potential fluctuations as well. Thus we might expect enhanced absorption over that calculated assuming a uniform distribution. The higher-lying state must have nodal properties that differ from the lower state, and this may affect the absorption as well. I know of no theoretical treatment that deals with these effects in any significant way.
5.4 Processes Contributing to TPA

We have described in great detail the theory of a single optical absorption event. In practice, in the TPA measurement, we measure the change in absorption resulting from a large number of excitations. Even for a single absorbed photon, several processes contribute to the change in absorption.

![Diagram showing processes contributing to TPA](image)

**Figure 5.1. Processes contributing to TPA.**

The absorption of an excitation photon, with energy $\hbar \omega_x$, lifts an electron from the valence band to the conduction band. It may be that the electron and hole remain correlated after the excitation, forming an exciton. This case cannot be represented on a one-electron diagram. The
TPA signal then corresponds to absorption by the exciton. The process of absorption by excitons has been observed by Robins and Kastner (1984), and will be discussed further in Chapter 13.

The more usual case in TPA measurements is that the electron and hole separate before the observation. After excitation, the electron and hole lose energy as rapidly as is allowed, since they have much more than $kT$ of excess energy; once they have lost enough energy to encounter localized states, however, further energy loss is much slower. At some time delay $t$ following their introduction, we measure the transmission of probe light, with photon energy $\hbar \omega_p$. Fig 5.1 shows the four processes that contribute to the absorption.

The most obvious process is absorption by the re-excitation of the excited electron back into the conduction band. The probe light generally must have an energy below the band gap (otherwise it would not be transmitted), but it may well have enough energy to re-excite a carrier, once the high-energy photons have done the initial excitation. Similarly, the excited hole can be re-excited to the valence band. These two processes both contribute to the TPA signal. Unfortunately, they are always present in equal numbers, so there is no way to separate the two signals in a homogeneous experiment, unless their spectral signatures can be unambiguously determined.

Two additional processes contribute to the TPA signal; these are the photo-induced transmission (transient photo-induced bleaching or TPB) processes. For example, it is no longer possible to excite an electron out of the state that the hole now occupies in the valence band. As a result, there is a decrease in absorption by that state. Recalling that the integrated absorption cross section cannot change (oscillator
strength sum rule), there must be just as much induced transmission as there is induced absorption. This bleaching will generally occur in different regions of the spectrum from the extra absorption, however: there is no bleaching of absorption unless there is absorption to begin with. For probe energies below the band gap, the TPB processes are not usually important. In the presence of long-lived sub-bandgap absorption the transient bleaching of this band may be important, as will see in Chapter 13.

5.5 Room-temperature TPA: Thermalizing Carriers

The room-temperature TPA spectrum, excited by 1.4 eV light, is shown in Fig. 5.2. As discussed in Chapter 2, the signal is plotted as a change in cross section Δσ per absorbed carrier. The data is shown at three different time delays, which differ by a factor of ten. The spectra are strikingly linear, rising from thresholds near 0.5eV, and continuing to rise up to 1.5eV, where the opacity of the sample prevents further measurement. The shape of the spectra will be discussed in detail in the next section.

One obvious feature of the data is the evolution of the spectral shape with time. For the plots shown, this can be best described as a linear shift to higher photon energies with the logarithmic increments of time. Such a logarithmic shift is precisely what is expected from the MT model, and indeed the observation of the spectral shift (Orenstein, Monroe, and Kastner, 1980) was a strong motivation in the development of the MT model by Orenstein and Kastner.

In order for the MT model to be applicable, the observed spectrum must be associated with the carriers undergoing multiple trapping, that is, the holes. The TPA signal results from photoexcitation of holes from
Figure 5.2. Room-temperature TPA spectrum, for three different time delays.

the band tails back into the valence band, as shown in Fig. 5.3. The threshold must correspond to excitation to the mobility edge or some similar feature in the valence band. Both the magnitude and the shift of the edge agree well with the value of $kT \ln(v_o t)$ at the relevant times, which are 0.43, 0.50, and 0.56 eV for $v_o = 2 \times 10^{12} \text{s}^{-1}$ and $t = 10 \mu\text{s}$, $100 \mu\text{s}$, and $1 \text{ms}$ respectively.

In the MT model, most of the thermalizing carriers are found at $E_d$; the density of carriers falls off exponentially above and below this energy (for $T < T_o$). Thus if the spectrum of the induced absorption from a
Figure 5.3. Postulated mechanism for room-temperature TPA: photo-ionization of thermalizing holes. (The figure shows the analogous process for electrons, to correspond to our discussion of MT.)

single carrier does not vary faster than this, the resultant spectrum ought to be equivalent to that resulting from absorption by a delta-function of carriers at $E_d$. The TPA experiment then reflects the product of a single DOS with the associated matrix element, rather than a convolution as is measured in normal optical absorption.

This conclusion is strengthened by the comparison of the data of Fig. 5.2 with experiments performed by Orenstein (1981). Orenstein used a higher excitation energy (~2.1eV), and the larger absorption coefficient gave a much higher excitation density, resulting in recombination. Nonetheless, although the magnitude of his data is decreasing with time, the shapes of the spectra, as well as the shift of the spectrum with
time, are quite similar. Since the number of shallow carriers is substantially reduced by recombination, relative to the number of carriers at $E_d$, the similarity shows that the latter carriers are responsible for the observed signal. In addition our observations vindicate Orenstein's procedure of normalizing the data to have the same magnitude at 1.4eV.

The association of the TPA signal with thermalizing carriers would seem to be further strengthened by the measurements of OKV in the presence of recombination. These authors measured the time dependence of the magnitude of the TPA at a fixed $\hbar \omega_p$ of 1.38eV. Since this energy is much greater than $kT ln(\nu_0 t)$, presumably such excitation can ionize all carriers with similar probability. In the absence of recombination there is no decay at this energy, so any spectral shifts are not substantially affecting the data. In the presence of recombination, the decay is proportional to $t^{-\alpha}$. All these effects are consistent with the predictions of MT, assuming that the 1.38eV TPA is a monitor of the total number of trapped carriers, independent of their energy. It should be stressed, however, that since the number of electrons is equal to the number of holes, the electrons could well be contributing to this signal as well; no recombination measurement could detect the difference. The observed spectra might well be the sum of an electron- and a hole-related part.

It is interesting that the thresholds agree so well with $kT ln(\nu_0 t)$, since this agreement seems to indicate the absence of significant electron-phonon coupling of the states. We have already mentioned the absence of such effects in the thermal excitation out of the band-tail states; their absence in optical excitation is an even stronger indication. The strength of this argument rests on an association of the threshold with the mobility edge, however, which in turn requires a
detailed understanding of the shape of the spectra. As discussed in the next section, such an understanding does not yet exist.

5.6 Shape of TPA Spectrum

The linear dependence of $\Delta \sigma$ on photon energy is wonderfully simple. Unfortunately, there seems to be no reason for the spectra to have such a dependence. According to Eqs. (5.3) to (5.5), the cross section is proportional to the density of final states and either the position matrix element squared \textit{times} $\hbar \omega$, or the momentum matrix element \textit{divided} by $\hbar \omega$. To determine the matrix-element-DOS product, then, one ought to plot $\Delta \sigma/\hbar \omega$ or $\Delta \sigma \hbar \omega$, depending on the matrix element desired. There seems to be no fundamental significance to $\Delta \sigma$ itself.

Figs. 5.4 and 5.5 show such plots. Obviously they show much greater curvature than the original plot, which may not be surprising since one does not expect a linear DOS. More importantly, the spectra at different time delays seem to change shape more than the $\Delta \sigma$ data, so that the simple interpretation of thermalizing carriers loses some of its simplicity, since the matrix elements must then depend on the initial energy. Of course the $\Delta \sigma$ data also changes shape, developing a low energy "foot" at longer times, but the slope of the linear part of the plots does not seem to change with time. The foot may be a result of the background effects discussed in Chapter 12 and 13, which become much more pronounced at long times; these room-temperature data really ought to be taken with a bleaching bias to check for these effects.

By associating the absorption with the photo-ionization of carriers thermalizing in band tails, we are finally able to apply the theoretical discussion of the early part of this chapter. Unfortunately, there are many uncertainties about the description of wavefunctions in disordered
Figure 5.4. Same as Fig. 5.2, but with $\Delta \sigma$ multiplied by $\tilde{\omega}_p$.

A discussion of the optical cross sections associated with states in the gap of crystalline semiconductors is given by Jaros (1977, 1980). For ionization of a highly localized state, the cross section has a threshold at the energy necessary to excite the carrier to the band edge, grows as the DOS in the band grows, but then drops off again at higher photon energies. The peak occurs roughly where the size of the wavefunction matches the wavelength of the final state; for higher ener-
gies the matrix element falls because of the cancellation between both positive and negative contributions to the integral. The peak generally no higher than twice the threshold. In the presence of electron-phonon coupling the entire spectrum is broadened, essentially by convolution with a Gaussian. This may wash out the threshold somewhat, but unless the broadening is quite large it will not totally obscure it.

This description is hard to reconcile with the observed spectrum. Gorenstein showed that the forms given by Jaros could simply not be made to fit the observed spectrum. The rise in the observed spectrum is too
gradual, and there is no evidence of a peak, even up to three times the threshold energy. There is room for some electron-phonon coupling in the spectrum, but not enough to dramatically change these conclusions.

An additional puzzle concerns the magnitude of the TPA signal. By doing measurements in the absence of recombination, we get an accurate measure of $\Delta \omega$ per absorbed photon. As shown in Fig. 5.2, this yields values of order $10^{-19}\text{cm}^2$. One would expect that if the induced absorption band is spread over a range of $\hbar \omega_p - 1\text{eV}$, the magnitude would be of order $10^{-16}\text{cm}^2$. It is conceivable, of course, that the quantum efficiency is very small, so that the contribution per surviving carrier has this magnitude, but the photocurrent quantum efficiency measurements (see Chapter 10) seem to contradict this possibility (recall that the current and the TPA should be sensitive to the same carriers). Although the discrepancy is not enormous, and may be accounted for by the the uncertainty inherent in the sum rule (Eq. 5.7) and the width of the spectrum, the small size of the TPA is a puzzle.

Further discussion of the spectra will be presented in Chapter 13, but here we mention three possible explanations of the peculiar shape of the spectra:

(1) Electrons are contributing to the spectrum. One certainly expects a contribution from electrons somewhere in the spectrum, although it could be at energies higher than the bandgap. We have already noted that the contributions from electrons and holes cannot be separated in a homogeneous measurements. Attempts to sweep out the holes and measure the residual spectrum (Thio and Kastner, private communication) proved unsuccessful. If the electrons are deeply trapped, with a threshold of, say, 1eV, then the sum of the two spectra might
yield something similar to that observed. Since the low-energy absorption would be dominated by the holes, the spectral shifts from thermalization would be as before. In this case the linearity of the spectrum is totally accidental. The cross section is still a puzzle.

Figure 5.6. Electron phonon-coupling model. For the form shown, there is little or no barrier to trapping, but a shift in the optical threshold equal to the depth of the trap.

(2) There is substantial electron-phonon coupling. We have already noted that the TPC results leave little room for a barrier to trapping. The configurational-coordinate diagram must then have the form illustrated
in Fig. 5.6. The spectral weight of the absorption would be shifted to higher photon energies. This would explain the small magnitude as well as the peculiar spectrum of the induced absorption. In this case the similarity of the threshold to $kT\ln(v_o t)$ would have to be more-or-less accidental. In addition, it is difficult to think of a mechanism in which the polaron is degenerate with the free hole at $Q=0$, when the binding energy is varying over a wide range.

(3) More exotic spectral effects. The calculations mentioned were for crystalline semiconductors. It was thought that since the optical absorption from a localized state depends primarily on the nodal properties of the wavefunction, the spectra would not be strongly affected by the disorder. It may be that the excitation from band-tail states to states near the mobility edge has special properties, for example the correlation effects mentioned earlier.

As discussed in Chapter 10, the optical properties of amorphous semiconductors in general are not very well understood. An understanding of the TPA spectrum may therefore be too much to hope for. In many ways, however, the TPA is a cleaner experiment than the optical absorption: it involves only a single DOS, rather than a convolution; there may be no excitonic effects to complicate the response; and even at a single temperature the thermalization of the carriers provides for excitations from several initial energies. Whether these features outweigh the complications, especially the multiple contributions to the spectrum, remains to be seen.
CHAPTER 6
LOW TEMPERATURE TRANSPORT ANOMALIES

The multiple-trapping model for dispersive transport has been tremendously successful. Even its simplest form, MTX, in which the DOS is exponential and the trapping coefficients independent of energy, suffices to explain a wide variety of observations. These observations include the recombination-free transport and both monomolecular and bimolecular recombination, as well as the peculiar form of the high temperature decays. At moderately low temperatures, however (below ~200K), important deviations from the MTX predictions are observed. In this chapter we will describe the observations, and the results of several tests of possible explanation. The next three chapters will be devoted to examining the most probable and exciting possibility, a transition to hopping-dominated thermalization, in greater theoretical detail.

6.1 Low temperature Photocurrent

We noted in previous chapters that the exponent of the time decay of the TPC provides reflects the energy dependence of the DOS. Specifically, if we assume constant capture coefficients,

$$\frac{\partial \ln i(t)}{\partial \ln t} = -1 + \frac{\partial \ln g(E)}{\partial E} \bigg|_{E=E_d}$$  \hspace{1cm} (6.1)

as long as the slope is not changing too rapidly. If the trapping coefficients change with energy, then $g(E)$ should be reinterpreted as the effective DOS, and $E$ as $-E_r$, as discussed in Section 4.1. The local variations are evaluated at the demarcation energy $E_d$, $kT \ln(\nu_0 t)$ below the transport energy. Thus measurements at different times and temperatures are sensitive to different parts of the DOS. For $\text{As}_2\text{Se}_3$ the
measurements imply an exponential DOS over depths from 0.3eV to 0.75eV, or four orders of magnitude in DOS.

Improvements in sensitivity made possible by the YAG laser (see Appendix A) and the Active Load Resistor (see Appendix E) made it possible to extend the measurements to substantially lower temperatures. As first reported by Monroe, Orenstein and Kastner (1981), the observations show dramatic deviations from the MTX predictions. The best results are collected in Fig 6.1, which shows the TFC decay for various temperatures. Note that the temperatures are much more closely spaced in the range 147 to 206K than elsewhere; as the temperature is lowered through this range the current begins to decay much more rapidly.

The transients in Fig. 6.1 are not power laws: they have a significantly smaller slope at long times, so they are concave upward. This observation, which required careful data acquisition, will help us to rule out some possible mechanisms for the transition. To exhibit the deviations from MTX most clearly, we have nonetheless force-fit these curves to single power laws; the results are shown in Fig. 6.2. The straight line shows the slope predicted by the MTX model; the value of $T_o = 580K$ was chosen to fit the high-T data.

There are many possible explanations for the change in behavior at low temperatures. For example, the microscopic trapping coefficients, characterizing multiphonon transition rates, may be exhibiting some temperature dependence as the phonon modes become depopulated. Alternatively, the multiple-trapping model itself may be breaking down, with transitions directly between the traps becoming important; we explore this possibility in following chapters. Because the hole current is decreasing with temperature, it may be that a previously unobserved
Figure 6.1. Transient Photocurrent at various temperatures. The lowest two transients used 1.7eV excitation; the higher temperatures were taken with $\hbar\omega_x = 1.18$eV.
Figure 6.2. Exponents of power-law fits to the data of Fig. 6.3. The error bars are derived from the deviations of the data from a straight-line fit.

electron current is showing up. The possibility of experimental artifacts that appear only at low temperatures must also be examined. The most straightforward possibility, however, is that multiple trapping is going on as before, but that the DOS close to the mobility edge, to which the low temperature measurements are sensitive, ceases to follow the same exponential dependence. We will examine each of these possibil-
ities, but since the original motivation of the experiment was to look for changes in the DOS we examine this possibility first.

6.2 DOS Interpretation

The interpretation of the low-temperature behavior as a DOS effect implies a DOS of the form shown in Fig 6.3: the steep slopes at low T imply a flatter DOS close to the mobility edge. Note, however, that since any variations in the DOS cause a deviation from one proportional to T, the sensitivity at low temperatures is poorer. For example, at T=0 the MT model predicts a $t^{-1}$ decay for any DOS.

A DOS of the form shown by the solid line helps to explain the low apparent microscopic mobility observed in As$_2$Se$_3$. This effective mobility $\mu_1$ differs from the true microscopic mobility $\mu_0$ by a factor $N_0/N_L$. Here $N_0$ is related to the true DOS at $E=0$, while $N_L$ is related to the extrapolated DOS, which for the form shown is larger by a factor of ~20. The observed value of $\mu_1=0.2 cm^2 V^{-1} s^{-1}$ then corresponds to $\mu_0 \approx 5$, which is quite consistent with expectations for extended state mobility at the mobility edge. In the next chapter, we will explore an alternative explanation for this low mobility, based on hopping.

Fortunately, we have a consistency check for the DOS explanation of the transition, which involves taking data at different temperatures, but changing the time at which the data is taken so that $kT \ln(\nu_0 t)$ remains constant; Fig. 6.4 shows such a comparison between data taken at 290K and at 185K. The figure shows the quantity $[t/(t_i)]^{-1}$ for various $kT \ln(\nu_0 t)$; this should be proportional to the DOS, except for the prefactor of Fig. 3.3. Since this prefactor depends on the slope, it will affect the shape of the final curve, but since the changes in slope for each temperature are small, this correction is minor. The data clearly
Figure 6.3 DOS required to explain the TPC data in the spectroscopic interpretation.

do not agree for the two temperatures. A pure DOS effect is thus not sufficient to explain the observations.

6.3 Electron-phonon coupling

One way to explain this problem is to introduce a temperature dependence to $v_o$, which shifts the curves horizontally; in the figure we used the value $v_o=2\times10^{12}\text{s}^{-1}$ deduced from time-of-flight data by OKV. The physical meaning of such an adjustment is that the multiphonon transition rates are beginning to change with temperature. The possible man-
Figure 6.4. Test of DOS model for low-T effects: comparison of inferred relative DOS \([i(t,t)]^{-1}\) for \(T=290\text{K}\) and \(185\text{K}\).

Manifestations of such changes are far more diverse than the simple horizontal shift, and could affect the shape of the spectrum in many other ways as well. I have no doubt that one could invent variations that reproduce the data. After all, even within the activated barrier approximation (see Section 4.1), there are two independent functions of energy \((g(E)\) and \(E_b(E))\) which can be adjusted to fit \(i(t,T)\).

One potential explanation in which the changes in the DOS are correlated with changes in capture coefficients involves a gap in the DOS near the mobility edge arising from electron-phonon coupling, as
suggested by Anderson (1972). The basic idea is that states deeper in the gap will be more localized and therefore experience a larger lowering of their energy as a result of electron-phonon coupling. Under certain conditions a gap will open in the DOS separating shallow, weakly-coupled states from deep, strongly-coupled states. We have already noted the lack of evidence for strong coupling in measurements at high temperatures but these may not be particularly relevant.

Because of the large degree of freedom, we cannot eliminate the possibility that multiphonon effects are responsible for the low-temperature anomalies in the TPC; such an explanation will have to be somewhat arbitrary, however, in the absence of more detailed information. In contrast, the hopping model introduced in the next chapter involves no new parameters, and predicts on the basis of known quantities that a transition must occur in the temperature range where we see these effects. Before discussing this model, we will examine various possible experimental artifacts that could contribute to the effect.

6.4 Background Effects

Several aspects of the experimental situation change as the temperature is lowered. It is therefore important to examine these changes to see if they may be responsible for the changes in time dependence.

The first type of potential artifact we will examine results from the buildup of a background of excitations at low temperatures. Such a buildup was described briefly in Section 3.10, and will be discussed in more detail in Chapter 12. In the latter discussion we estimate that the number of carriers in the background is larger than the number introduced per pulse by the ratio of the lifetime to the repetition time; since the lifetime becomes very long at low temperatures this
ratio may be very large.

Two potential manifestations of a large background are saturation and recombination. Saturation of the band tails would be expected to make the decay slower, and cannot explain the observations. Recombination, by contrast, should speed up the decay. This decay would be monomolecular in the sense that the recombination would not immediately affect the density of recombination centers; the decay should therefore be proportional to $t^{-1-\alpha}$. The recombination should be most important at long times, that is the log-log plots should be concave downward (steeper at long times). In fact, the opposite effect is observed; the possibility of testing for this mechanism was a major reason for taking the data of Fig. 6.1 very carefully. We may therefore conclude that the presence of a background will not produce the observed effects, unless additional deviations from the MTX model are introduced.

6.5 Space Charge

Another possible artifact, which is similar to the background effects, involves buildup of space charge in the sample. This is quite plausible, since at 150K the dark conductivity is around $10^{-20}\Omega^{-1}cm^{-1}$, while the conductivity of the region illuminated by the laser is substantially higher. As discussed in detail in Appendix B, inhomogeneous conductivity results in charging effects which screen the voltage falling across the more conductive region. In principle, measuring the conductivity with an ac field will eliminate this problem, but while this technique worked at room temperature, the available signal-to-noise ratio was insufficient to extend this technique to low temperatures, mainly because a low-voltage ac source was used.
It is important to note, however, that space-charge buildup will only affect the magnitude of the current; indeed, I consider the relative magnitudes of the data taken at the various low temperatures to be rather unreliable for precisely this reason, and have focussed primarily on the time dependence. This time dependence will be unchanged by the charging effects as long as the space charge does not rearrange on the time scale of the measurement (see Appendix B).

Such rearrangement will in fact be negligible, because the distance moved by the charges is very small: even at room temperature, where $\alpha \sim 0.55$, using the values $\mu_1 = \mu_0 N_o/N_L = 0.2 \text{cm}^2 \text{V}^{-1} \text{s}^{-1}$, $F = 2000 \text{V/0.5cm}$, and $v_o = 2 \times 10^{12} \text{s}^{-1}$, the carriers move only a distance $l = \int F\mu(t)dt - F\mu_1/v_o \alpha (v_o \alpha)^{-\alpha}$. This amounts to only $10^{-3} \text{cm}$ in the 0.1s time between pulses. At low temperatures they travel an even smaller distance; it is therefore inconceivable that they can move between regions with substantially different fields, except perhaps very close to the contacts. [Some effects of focusing the light close to the contacts were observed, resulting in a change in magnitude of a factor of five in the presence of steady-state HeNe illumination, but no change in the time dependence.]

The concerted motion of all the carriers could in principle change the fields because of the accumulation of space charge, however. This effect is reduced because there are overlapping positive and negative charge distributions created by the pulse, so that free charge develops only to the extent that the distributions separate. The total charge in a distribution of width $w$ must exceed that generating the field, $CV$, by a factor $w/l$ in order for the nonoverlapping charge to be comparable to $CV$. For $C=10 \mu\text{F}$, $V=2000 \text{V}$, and $l$ given above, we need a few microjoules in
a mm-size spot to see the charging at room temperature by 0.1s. In other words, this condition may be marginal and should be checked at room temperature. At low temperatures, however, there should be no problem, and this cannot cause the change in time dependence. More importantly, like recombination it would give the wrong curvature.

Of course the true test for average-power effects is to change the average power and see if the results change. Although this is difficult at low temperatures, I performed several checks for linearity with voltage, with pulse energy, and for effects caused by the extra illumination. Although changes in magnitude were observed (suggesting some charging), the time dependence continued to be anomalously steep. It can therefore be concluded that the effects described were not responsible for the anomalous time dependence.

6.6 Photon Energy

The initial results demonstrating the low-temperature anomalies were obtained using the low photon energy fundamental of the YAG laser ($\hbar\omega_x = 1.17$eV), as were the careful studies of the crossover regime. This excitation gives room temperature photocurrent transients that are similar to those obtained with high-energy light, but there are small differences, as described in Chapter 10. It is therefore important to be sure that the low-temperature anomalies are also seen with high-energy excitation.

The answer seems to be that they are. An example is the T=77K decay shown in Fig. 6.1, which was taken with $\hbar\omega_x = 1.7$eV. The temperature is low, so within MTX one expects the slope to be large, but the observed slope is even larger. The anomalies are therefore not limited to small excitation energies. Nonetheless, it would be interesting to study the
crossover region with two different photon energies. The differences to be discussed in Chapter 10 might become more dramatic before the eventual \( t^{-1} \) decay is reached. Such an observation would provide information both about the nature of the temperature transition and the nature of the final state for various excitation energies.

6.7 Short-time Measurements

Although it was not appreciated at the time, the anomalous transport at low temperatures was first observed in short-time measurements. The integrated photocurrent during the laser pulse was measured:

\[
Q_p = \int_{\text{during pulse}} i(t) \, dt \tag{6.2}
\]

which is roughly equivalent to

\[
Q_p = \int_0^{t_p} i_o(t) \, dt \tag{6.3}
\]

in which \( t_p \) is the length of the pulse and \( i_o(t) \) is the response to a delta-function with the same pulse energy. Thus the peak current measurement provides an indication of the decay for times shorter than the pulse length.

If the current follows the MTX form, then \( Q_p \) should go as

\[
Q_p \bigg|_{MT} = \int_0^{t_p} (\nu_o t)^{-1+T/T_o} \, dt \tag{6.4}
\]

\[
= \frac{T}{\nu_o t_p} \frac{T}{T_o} \tag{6.5}
\]

\[
= \frac{T}{T_o} \exp \left[ \frac{T}{T_o / \ln(\nu_o t_p)} \right] \tag{6.6}
\]

The data, shown in Fig 6.5, are roughly exponentially dependent on temperature, but at low temperatures the data become significantly less dependent on temperature. These data are consistent with a crossover to
a $t^{-1}$ decay whose magnitude depends only weakly on temperature.

![Graph](image)

**Figure 6.5.** Integrated prompt photocurrent. The data shows the total charge per incident photon flowing during the 10ns-long laser pulse at different temperatures.

### 6.8 Zero-voltage Transients

During the course of the low-temperature photocurrent measurements, a very peculiar phenomenon was observed—so peculiar that it received the appellation "Item X" in my lab notebook to avoid prejudicing the interpretation. No physical mechanism has been deduced for this phenomenon, so this section simply describes some of its characteris-
tics.

The phenomenon, which will be referred to as the zero voltage transient, or ZVT, is a current transient induced by the laser pulse: blocking the laser stops the effect. It does not, however, require that any voltage be applied to the sample. This is its peculiar characteristic, since without voltage there ought to be no preference for current flow in any particular direction. Any observed current must result from asymmetry, either in the illumination, or in the sample itself.

The ZVT was first noticed when doing photocurrent measurements with the YAG laser (\(\hbar\omega_x = 1.17\text{eV}\)) at low temperature (the following characterization was done at 77K). [The direct output of the YAG laser has significant beam inhomogeneity; at the ~4m distance at which these experiments were done, however, I think the inhomogeneity is largely gone.] Individual transients fluctuated in sign from pulse to pulse, but the averaged signal had definite polarity. A sample of the averaged data is shown in Fig. 6.6; it consists of a ~1\(\mu\text{s}\)-long peak followed by an oscillatory part.

For comparison, the photocurrent response with ~100V applied to the 5mm sample is also shown. The ZVT seems to be reproducible enough that subsequent averaged signals subtract to zero, although there are some indications of nonreproducibility from day to day, which may be related to changes in the geometry. Note also that the ZVT decays more slowly than the current, being comparable at short times but much larger at long times. Prior to the ZVT shown, 2000V had been applied to the sample, but there is little polarization left by the time of the measurement. The different time dependence of the ZVT also shows that it is not a simple polarization effect, and this was verified by prolonged zero-
voltage resting.

At room temperature a signal was also observed, which was of the same order of magnitude in size and therefore much smaller relative to the photocurrent. The time dependence was also different, and as shown in Fig 6.7, changes under continual illumination. The changes suggest that there are two processes, one which depends on built-up charge density and one that does not.

Not enough work was done to characterize this phenomenon, since at the time it was deemed less interesting than other possible experiments. Nonetheless, it is quite puzzling, and its surprising nature can be deduced by the following observation: Suppose that the current arises from local inhomogeneity in the sample causing a preference for current
in some direction; suppose further that this preferred direction varies randomly on a length scale $\xi$. The volume over which this effect is averaged is several $\text{mm}^3$; any random effect will therefore be reduced relative to a uniform effect by roughly $(\xi/0.1\text{cm})^{3/2}$. We see an effect comparable to a uniform field of order $10^3 \text{V/cm}$. Even if $\xi$ is as large as 1$\mu$m, the local effect corresponds to a field of order $10^7 \text{V/cm}$, a very large field indeed.

Of course it is conceivable that the sample is inhomogeneous on an even longer length scale, although there is no evidence for this in other experiments such as optical transmission; such inhomogeneity would
have disturbing implications for other measurements. It seems more likely that the ZVT's arise from a totally different effect. This may have to do with the peculiar nature of the transport at low temperatures, or it may be totally unrelated, for example, a photo-acoustic effect. It remains an intriguing puzzle.

6.9 Low-temperature TPA

One of the most important pieces of evidence in the low-temperature behavior was already on hand when Joe Orenstein completed his thesis in 1981: the low temperature evolution of the TPA spectrum. As discussed in the preceding chapter, the room-temperature TPA spectrum provides the most direct evidence of the thermalization of carriers associated with MT. One sees a shift in the spectrum to higher photon energies with a magnitude that is consistent with carriers being photo-excited from an energy $kT\ln(v_o t)$.

At 20K, the thermalization due to MT should be negligible, less than one tenth that at room temperature. Instead, as shown in Fig. 6.8, the spectra still appear to shift, approximately half as fast as at room temperature. It should be noted that the data are really decaying in magnitude, and have thus been normalized to have equal magnitude at 1.4eV; this procedure has been proved accurate at room temperature, but not at low temperatures. In addition, contributions other than the thermalizing carriers appear at low temperatures, as we discuss in the last part of this thesis; these undoubtedly contribute to the markedly different shape of the low-temperature spectra.

Nonetheless, taken at face value, these spectra suggest that carriers are thermalizing much more rapidly than their thermal excitation to the mobility edge would allow. An obvious possibility is that the
Figure 6.8. TPA spectra at 20K, from Orenstein (1981).

carriers are hopping directly between localized states, and this is the most probable explanation of the low-temperature anomalies. The development of a model for such a process is the major accomplishment in this thesis, and this model is presented in the next three chapters. According to this model, the alternative thermalization process must begin at temperatures comparable to those of the observed transition. Together with the absence of viable alternative models, as discussed in this chapter, I believe this constitutes strong evidence for the hopping model, which we will now examine.
CHAPTER 7
SIMPLE DESCRIPTION OF HOPPING MODEL

The preceding chapter gave experimental evidence for dramatic new behavior of the TPC and TPA at low temperatures. In this chapter we explore one explanation of these observations in detail: a transition from thermally-activated energy relaxation as in MT to hopping-dominated relaxation as the temperature is lowered. We develop a simple picture of the basic features of each of the two regimes of energy relaxation, which is consistent with experimental observations. We also give a simple estimate of the "segregation time" separating the two regimes, as a function of temperature. A more complete and mathematical description is given in the next chapter. The ideas presented here have important implications for transport in general, even at high temperatures, and in Chapter 9 we discuss some of the observations on which it sheds light.

7.1 Overview of Hopping Model

The critical difference between the two relaxation mechanisms is whether the rate-limiting step in the relaxation involves transitions upward or downward in energy, as illustrated in Fig. 7.1. Clearly the hopping-down mechanism will become more important as the thermal energy available for excitation to higher energies becomes smaller. Indeed, hopping has long played an important role in the understanding of the low temperature behavior of the dc transport properties of disordered materials (Mott and Davis, 1979). The well-known variable-range hopping (VRH) process (Mott, 1968), however, involves hopping near the Fermi level, while the present discussion involves carriers in the band tails.
Fig. 7.1 Illustration of the two mechanisms for energy relaxation (a) Hopping Down: carriers jump directly to lower energy states. (b) Hopping Up: carriers are thermally activated to higher energies, where the DOS is higher, undergo transitions within these "transport states," and are subsequently retrapped at lower energies.

As before, we describe the redistribution of only one type of carrier, electrons, which move within the exponential DOS deduced from the MT model, Eq (3.17). In MT, the carriers move to deeper states only by thermal activation to the mobility edge and subsequent retrapping in the deeper states. In the present discussion, we also include hopping, that is, transitions directly between the various localized states. Although
hopping is always possible, the low DOS (but not the small matrix elements, see Section 5.3) typical of localized-to-localized transitions reduce the rates relative to localized-to-extended transitions. Nonetheless, since carriers require less energy to hop to another localized state than to be thermally excited to the mobility edge, hopping transitions provide an important mechanism for carrier redistribution at lower temperatures. Although the exponential DOS was deduced without taking hopping into account, we will find that both the high and low temperature behavior of the TPC and TPA are consistent with hopping relaxation in an exponential DOS.

Rigorous analysis of the hopping problem is vastly more complicated than that for MT (Grünewald, Pohlmann, Movaghar, and Würtz, 1984; and Butcher, McInnes and Summerfield, 1983, and references therein). In contrast, we shall see that the intuitive picture (Orenstein and Kastner, 1981; Tiedje and Rose, 1981) developed for MT is useful for the hopping case as well. The reason for the success of the simple approach is the same in both cases: the tremendously broad distribution of rates for various types of centers, which implies that the occupation of most centers is determined by rates that are either very fast or very slow, compared to the relevant time scale. For the hopping case, we must also address the rather subtle issue of how to determine the appropriate rate, as discussed in the next chapter.

For pulsed excitation, the important time scale is time delay since the introduction of the carriers, because processes faster than this will already have played themselves out. In both MT and hopping, fast states have a steady-state occupation, reflecting a balance between carriers entering and leaving, while slow states have an occupancy equal to
the total number of carriers that have hopped to them. The critical aspects of the redistribution are determined by those states that have a release time comparable to the time delay; these marginal states will reflect the relaxation mechanism.

7.2 Microscopics of Hopping for Various Energies

Before discussing the overall redistribution of carriers, we must discuss the microscopics of the hopping process between a pair of localized states. Since our primary goal is to explore the general features of the hopping mechanism, we will assume the simplest form for the hopping rate consistent with the restriction imposed by detailed balance that the hopping rates up and down in energy differ by a Boltzmann factor:

\[ v_{if} = v_o \exp(-2\gamma R_{if}) \times \begin{cases} \exp(-\Delta E/kT) & \Delta E > 0 \\ 1 & \Delta E < 0 \end{cases}. \] (7.1)

Here \( v_{if} \) is the hopping rate between occupied initial site \( i \) and unoccupied final site \( f \), which are separated in distance by \( R_{if} \) and in energy by \( \Delta E = E_f - E_i \). \( v_o \) is an attempt-to-hop frequency and is expected to be of the order of phonon frequencies, \(-10^{12} \text{s}^{-1}\), and \( \gamma^{-1} \) is the decay length for the wavefunctions. This form is discussed in more detail in section 8.11.

The hopping rate has two competing exponential factors: one in energy, and one in distance reflecting wavefunction overlap; the outcome of their competition depends on the initial energy: Carriers in shallow states will hop rapidly to states nearby in energy, but carriers in deep states overlap very little with with other deep states, and thermal excitation to shallower states will be faster. The energy separating these two types of state, the (temperature-dependent) transport energy
$E_t$, plays an important role throughout our discussion.

$E_t$ can be estimated by maximizing the hopping rate, that is, the product $\exp(-E/kT) \exp(-2\gamma R)$, or, more conveniently, its logarithm. The logarithmic derivative of the Boltzmann factor with respect to energy is just $-1/kT$. The overlap factor depends on the typical distance one must go to find a state at the final energy in question, which will be $R_{\text{typ}} = C N_L^{-1/3} \exp(-E/3kT_0)$, where $C$ is a numerical factor of order unity. The logarithmic derivative of $\exp(-2\gamma R_{\text{typ}})$ is then $C N_L^{-1/3}/3kT_0 \exp(-E/3kT_0)$. Setting the magnitudes of the two expressions equal, we find that the overlap is the more important factor for energies greater than

$$E_t = E_0 - 3kT_0 \ln[T_0/T], \quad (7.2)$$

in which

$$E_0 = kT_0 \ln \left[ \frac{nN_L}{6\gamma^3 kT_0} \right]. \quad (7.3)$$

In determining the numerical factors defining $E_0$ we have used the results derived in the next chapter, rather than the heuristic derivation above. We expect that $E_0$ is not very far from the mobility edge, since the localized wavefunctions should just overlap at the mobility edge. The second term of Eq. (7.2) reflects the competition between the Boltzmann factor and the typical overlap, and causes the transport energy to become dramatically deeper as the temperature is lowered. For example, in $\text{As}_2\text{Se}_3$, $3kT_0$ is approximately 0.15eV, so this shift is of order 0.1eV at room temperature; it can be much larger at low temperatures.

In the variable-range hopping problem, the carriers hop shorter
distances at high temperatures, because of the larger density of energetically accessible sites. A similar effect occurs in the present case, except that when the carriers begin to hop up by more than $kT_0$ a catastrophic change occurs because of the exponential increase in the DOS; carriers will then hop all the way up to $E_t$. At $E_t$ further increases in the wavefunction overlap no longer compensate for the Boltzmann factor. The competition which determines $E_t$ does not depend on the initial energy: all carriers in deep states hop to the same set of final states.

Another difference between the present situation and VRH is that for a transient experiment carriers can hop down in energy, since there is no quasi-Fermi level below which traps are filled. This requires little or no thermal activation, but will be slow if the DOS is small. The competition between thermal activation to an exponentially higher DOS and hopping down results in two regimes of redistribution dominated by the two processes illustrated in Fig. 7.1. Because of the exponential DOS, these processes are rather distinct.

7.3 Energy Relaxation

The typical hopping rate at a given energy depends on which mechanism (hopping up or hopping down) is faster for carriers at that energy. Whatever the mechanism, the typical rate falls very rapidly as the energy of the initial state becomes deeper: For shallow states the rate is determined by hopping to lower states, so the decrease of the density of states causes the decrease of the rate. For deeper states the rate is limited by thermal activation to $E_t$, so the Boltzmann factor slows the rate.

This rapid decrease means that carriers in shallow states typically hop much faster than those in deep states. The division of states into
fast and slow is thus roughly equivalent to a division of states in energy, as it was in MT. This means we can again define a demarcation energy $E_d$ dividing fast and slow states, at which the typical total leaving rate $v_{\text{typ}}$ is comparable to the time delay since the carriers were introduced ($v_{\text{typ}}(E_d)t - 1$). States shallower than $E_d$ have fast rates, and carriers leave as fast as they arrive. They accumulate in states deeper than $E_d$, which have slow release rates and still contain all carriers that have hopped into them.

The critical aspects of the redistribution of carriers are determined by those carriers trapped at the marginal states with a release time $v_{\text{typ}}^{-1}$ comparable to the time delay, that is, those near $E_d(t)$. Because of the rapidly decreasing DOS, carriers that are released from these states will most likely become immobilized again at an energy not far below $E_d$, since there are more states there than at deeper energies. Thus, at any time most of the carriers are near $E_d$.

The overall behavior of the energy relaxation depends on the process (hopping up or hopping down) by which the marginal states release their carriers. That is, it depends on whether the demarcation energy is shallower or deeper than the transport energy. Because the demarcation energy moves deeper with time, the two mechanisms of redistribution will be active in different time regimes, with hopping down governing redistribution at short times, when $E_d > E_c$, and thermal activation governing at long times. We refer to the time of the crossover as the segregation time, $t_s$, for reasons that will be explained below.

### 7.4 Hopping-Down Regime

At early times, when the demarcation energy is shallow, the dominant mechanism of carrier redistribution is hopping to lower energies,
since the high DOS makes this process faster than thermal activation. Since thermal energy is not necessary by this process, we can gain insight into the redistribution by assuming that carriers can only hop down in energy. The number of states available for hopping is just the total number lower in energy than the initial energy $E_i$ (which is $N_L \exp(E_i/kT_0)$), so the typical rate is the rate for hopping the typical distance to a lower state:

$$v_{typ}(E_i) = v_0 e^{-6 \exp[-e^{-E_i/3kT_0}]} \quad E_i > E_t(T) \quad (7.4)$$

and the demarcation energy is

$$E_d(t) = E_0 - 3kT_0 \ln 6 - 3kT_0 \ln[\ln(v_0 t)] \quad t << \tau_s(T) \quad (7.5)$$

where the $\ln 6$ again is calculated in the next chapter.

This form was first derived by Kastner (1980) to describe the spectral shifts of luminescence at low temperatures. It shows that the carriers will lose energy rapidly at first, but their energy loss slows markedly (even on a logarithmic scale) as they encounter the lower DOS at deeper energies.

Although the process of energy relaxation, hopping down directly to other localized states, is different from the thermal activation relevant to the MT case, the distribution of carriers is similar, since it results from the distinction between fast and slow on the basis of their energy: states with $E < E_d$ will have an occupation determined by carriers hopping down from higher energies, while states with $E > E_d$ will be occupied only to the extent that carriers hop up from below.

### 7.5 Transient Photocurrent

The TPC observed during the hopping down process is associated with the thermalization process itself, as opposed to the MT case in which
the current simply reflects the occupation of the transport states. Because the shallow states are not in equilibrium, the current cannot be determined by the quasi-equilibrium argument used for MT in chapter 3. An alternative derivation of the current works equally well for both cases, however, as we now show.

Whatever the mechanism, the rate-limiting step for the redistribution of carriers is their release from marginal states, which have leaving rates of order 1/t. The rapidly decreasing DOS means that that a substantial fraction of the carriers will indeed be trapped in marginal states, so the number of carriers per unit time that become free to move will be a multiple of order 1/t of the total number of carriers introduced.

The current is the product of the rate of release of carriers (≈ 1/t) and the amount of current they carry before becoming immobilized again. For MT, carriers carry current as long as they are in the transport states or in shallow traps, until finally they become trapped below $E_d$. Since the number of deep states is falling as $t^{-\alpha}$ (recall $\alpha = T/T_\alpha$) and the lifetime is inversely proportional to the number of deep states, the TPC is just a proportional to $t^{-1}/t^{-\alpha} = t^{-1+\alpha}$.

In the hopping-down regime, by contrast, once a carrier hops it immediately finds itself in a substantially lower DOS, and is thus immobilized. As a result the current associated with the thermalization process will be roughly proportional to 1/t. Thus the hopping-down model provides a reasonable explanation of the steep (roughly 1/t) power-laws observed at low temperatures.
7.6 Segregation Time

As the carriers continue to lose energy by hopping down, they eventually reach the transport energy for the temperature in question. Naturally, since the transport energy is deeper at lower temperatures, the time at which this occurs is longer. This delay is enhanced by the fact that the energy relaxation gets much slower as the carriers become deeper. The time at which the temperature independent demarcation energy of Eq. (7.5) reaches the transport energy of Eq. (7.2) is the segregation time

\[ \tau_s = \nu_0^{-1} \exp(3T_0/T) \]  

A more physical derivation involves setting the thermalization rates equal; that is, asking: which process allows the carriers to get deeper faster? The rate of change of \( E_d \) with respect to \( \ln t \) in the hopping-down regime (Eq. 7.5), is \(-3kT_0/\ln(\nu_0 t)\). This becomes slower than the equivalent quantity for MT (-kT) at a time \( \tau_s \).

The segregation time \( \tau_s \) determines the dominant energy relaxation process for a given experiment: For \( t \ll \tau_s \), the behavior is dominated by downward transitions directly between localized states. For \( t \gg \tau_s \), the behavior is governed by thermal activation. Similarly, for experiments with a fixed time scale, only low-temperature measurements will be directly affected by hopping.

For \( t \gg \tau_s \), the shallow states are in internal equilibrium, and the redistribution is limited by thermal excitation from deep states to \( E_t \). As first pointed out by Schmidlin (1980), we can distinguish between traps (which do not communicate with one another) and the transport states (for which transitions to other transport states dominate). The new feature of the present work is that for an exponential DOS, this
distinction is equivalent to a distinction between states deeper and shallower than $E_t$. $\tau_s$ is the time at which the transport states come into equilibrium. It thus represents the time at which a distinction between traps and transport states becomes meaningful, that is the time at which the system becomes segregated into the two types of state.

In addition to its role in separating the two regimes of carrier thermalization, the segregation time has microscopic significance. Since $\tau_s$ is the time delay at which $E_d$ reaches $E_t$, and since $E_t$ is defined by having a typical rate equal to the inverse of the time delay, $\tau_s$ is the typical hopping time for carriers at $E_t$. Since the internal transitions of the bath of transport states determine its response to external fields as well as the rate of transitions into and out of the bath, $\tau_s$ is an important parameter microscopically as well as macroscopically. We will discuss the possibility of measuring this effect in ac conductivity in Section 9.5.

7.7 Hopping-up (Multiple-Trapping) Regime

For all times longer than the segregation time, the carriers limiting the redistribution are in a low DOS, so hopping to deeper states is very slow. It is faster for a carrier to be thermally excited to shallower states, at which the high DOS makes it much easier to find a close center. For $t >> \tau_s$, when the carriers will have gotten much deeper than $E_t$, they will primarily hop back up to states near $E_t$. Since the competition is independent of initial energy and the set of final states is the same, the only contribution to the rate that depends on the initial energy is the Boltzmann factor. Thus the rate of hopping up will have a simple activated form:

$$v_{typ}(E_i) = \tau_s^{-1}(T) e^{-\frac{(E_t-E_i)}{kT}} \quad E_i < E_t(T), \quad (7.7)$$
if the hopping rate has the simple form of Eq. (7.1).

This form is identical to that assumed for MTX, except that the carriers are now being excited to $E_t$ rather than to the mobility edge. In addition, since the typical overlap is smaller at the transport energy than at higher energies, the prefactor is reduced to $\tau_s^{-1}$, which is the typical hopping rate at $E_t$. Note that in the hopping problem there is a close relationship between the transitions responsible for thermal activation, which determine the prefactor in Eq. (7.7), and those responsible for transport, which determine the transfer rate at $E_t$.

The carrier redistribution is determined by carriers near the demarcation energy, which is, inverting Eq. (7.7),

$$E_d(t) = E_t - kT \ln(t/\tau_s) \quad t \gg \tau_s$$

(7.8)

This is a very important result: when we allow carriers to hop in a rapidly decreasing DOS, their redistribution at long times is governed by processes very similar to those that determine the redistribution if only localized-to-extended transitions are allowed, except that the time scale and the zero of energy have been changed. This similarity between hopping and multiple trapping in an exponential DOS was shown in numerical simulations by Silver, Schoenherr and Baessler (1982); it is a consequence of the fact that, once $E_d < E_t$, carriers move primarily by thermal excitation to higher states. This may be particularly true for the system simulated by Silver et al.: their system was dilute ($N_L/\gamma^{3.01}$, for our definition of $\gamma$), which would be expected to suppress the hopping down process. This may explain why the data that they showed in their paper do not exhibit a hopping down regime.
Because of the similarity of the hopping-up regime to MT, we can directly transfer many of the known facts about the latter problem. For example, the occupation of states in the hopping up regime is similar to that for MT: deep states (below $E_c$) have a constant occupancy, while the shallower states are occupied according to a Boltzmann distribution. The shallow states are in quasi-equilibrium because the rate of exchange of carriers between the shallow states is much faster than the rate of trapping into the deep states.

A rough estimate of the magnitude of the current can be made as follows: The main temperature dependence of the mobility comes from the variation of the typical hopping time at $E_c$, $\tau_s^{-1}$. The effective degeneracy factor $N_c/N_L$ for the hopping process ought to be of order unity for temperatures somewhat lower than $T_o$, since the DOS is continuous through $E_c$. The current should therefore be proportional to

$$i(t) \propto (\nu_o \tau_s)^{-1} (t/\tau_s)^{-1+\alpha} \tag{7.9}$$

$$= (\nu_o \tau_s)^{-\alpha} (\nu_o t)^{-1+\alpha}$$

But according to Eq. (7.6),

$$\left(\nu_o \tau_s\right)^{-\alpha} = \varepsilon^{-3}, \tag{7.10}$$

independent of temperature; the power-law transients in the MT regime should still roughly focus on a double logarithmic plot, at a time of order $\nu_o^{-1}$. Thus both the time dependence and the temperature dependence of the TPC for $t >> \tau_s$ will be essentially the same, whether conduction is by hopping or by extended states. It is interesting to note, however, that the magnitude of the effective mobility to which such plots focus, $\nu_o N_c/N_L \sim 0.2 \text{cm}^2\text{V}^{-1}\text{s}^{-1}$, (Pfister and Scher, 1978, Orenstein et al 1982) would be rather too small for an extended state mobility (assuming
\( N_c/N_L - 1 \), see Section 6.2), but quite reasonable for hopping. We will discuss the magnitude of the current in greater detail in Section 8.8.

It should be apparent by this point that transport near the transport energy has many of the properties that are traditionally attributed to extended states. In fact, because of the exponential DOS, many experiments give the same results whether or not hopping is important. Nonetheless, there are important differences, as we discuss in Chapter 10.

7.8 Comparison with Transient Experiments

The present picture is very successful in describing the transient photocurrent data of Fig. 6.1, explaining (1) the MT-like power law dependence at high temperatures, (2) the more rapid decay at low temperatures, and (3) the concave upward character of the curves in the transition region. This last observation shows that the rapid decay regime occurs at short times, and that the crossover moves to longer times as the temperature is lowered, just as the model predicts. If we use a value for \( v_o \) of \( 10^{12} \text{s}^{-1} \), and \( T_o = 550 \text{K} \), we predict that the transition should be observed in the time range of Fig. 6.1 at temperatures around 120K. Considering the simplicity of the model, this agreement is quite good.

We do not have a prediction of the form of the current in the crossover regime. For amusement, in Fig 7.3 we have fit the data to the simple form

\[
i(t) = \frac{A}{t} + B(T) t^{-1+T/T_o}, \tag{7.11}
\]

with the same value of \( A \) and \( T_o \) for all the curves. The agreement is reasonable.
Figure 7.2. Predictions of Hopping Model for TPC. At short times, the current decays as $t^{-1}$ by a hopping-down mechanism, which rolls over to MT-behavior at $\tau_s$; $\tau_s$ becomes longer at lower temperatures.

We also note that the energy relaxation due to the hopping-down process is in agreement with the observed spectral shifts in the TPA. We associate this shift with the energy relaxation of carriers hopping down in energy; indeed, the TPA experiment provides the most direct evidence for excess energy relaxation.
Figure 7.3. Fits to Photocurrent Data. The data of Fig. 6.1 are fit with the sum of a $1/t$ part which is independent of temperature and a temperature dependent MT-like power law.
CHAPTER 8

MATHEMATICAL DESCRIPTION OF HOPPING

In this chapter we provide a more substantive mathematical basis for the description given in the previous chapter, by including the distributions of various quantities as a function of energy. This approach, which also provides powerful insight into the physics, will be introduced in the context of a simple analysis of the hopping-down process. We then include the possibility of thermal activation, and estimate the typical total hopping rate for carriers at various energies. Using this estimate we will derive the $T^{1/2}$ law for variable range hopping in a simple way; then we will apply this procedure to energy relaxation in an exponential DOS. Finally, we calculate the magnitude of the current in the MT regime in preparation for our treatment in the next chapter of the dc transport properties of exponential band tails. Our approach is closely related to the description given by Grünewald and Pohlmann (1979).

8.1 Typical Rate Approach

The complications introduced by the inclusion of hopping are formidable: In MT, the occupation of each state can be determined directly from a knowledge of the past and present occupation of the extended states alone. In the hopping case, a carrier in any state can in principle hop to any other state, so the occupancy of each state is intimately related to the (statistically distributed) distances and energies of nearby states, and to their occupancy. A rigorous analysis therefore requires a careful self-consistent treatment of the occupancies of all
states; see, for example Grünwald et al., 1984, and Butcher et al., 1983. Insuring self-consistency is difficult because an anomalously fast rate will cause the occupancy of nearby states to be anomalous as well. The "typical rate" procedure is designed to avoid this problem by ignoring **all** anomalous processes. This is of course the intent of the more rigorous work as well; the problem lies in determining what constitutes anomalous behavior.

The exponential dependence of the hopping rate on distance gives causes a tremendous variation in the hopping rates, even among centers with the same energy. Since the total rate at which carriers hop away is dominated by the fastest available hop, a center with another center unusually close will have a total leaving rate that is exponentially faster than most other centers with the same energy. It is critical to note, however, that this anomalously fast center is not very important to the overall redistribution: a carrier placed on such a center will hop away rapidly, but no more hops will occur from that center until a new carrier is supplied to it. Since it is unlikely that the supply rate is also anomalously fast, the total flux of carriers through the state will be not be nearly as large as its anomalous rate would suggest. This is illustrated in Fig. 8.1.

These considerations lead us to emphasize not the **average** hopping rate (which is dominated by the rare, exponentially large rates) but a **typical** hopping rate, which is a rate which will be found with some reasonable probability. If we take that probability to be one half, we see that the median rate is a typical rate. However, if the use of a typical rate is legitimate, the results should be insensitive to the exact procedure for choosing the typical rate. In this chapter we will use the
Figure 8.1. Unimportance of anomalously fast hops. If one out of ten centers (top) has a rate 100 times that of its fellows, the average rate will be ten times that rate, but the number of remaining carriers (bottom) will remain large until a more typical decay time, unless exchange between centers is faster.

"fastest expectable rate" criterion described in the next section.

8.2 Fastest Expectable Rate

In the hopping problem there is a tremendous variety of local environments for centers, so the hopping rates $v_{ij}$ between pairs of centers are spread over many orders of magnitude. It is therefore useful to specify their distribution in the logarithm of the rate, normalized to the fastest rate $v_0$: $\ln(v_0/v)$. We define $N[\ln(v_0/v)]d\ln v$ as the expected number of sites to which a carrier can hop with a rate between
\( v \) and \( ve^{\Delta \ln v} \). An example of such a distribution, derived in the next section, is shown as the solid line in Fig. 8.2.

![Diagram showing the distribution of rates with density of channels, total density, and fastest expectable rate labeled.]

Figure 8.2. Distribution of Rates. The solid line is the density of individual rates, the dashed line is the distribution of total rates.

Suppose that a carrier is placed on a site \( i \) at time \( t = 0 \). The probability that it is still there at time \( t \) is simply \( \exp(-v_{\text{tot}}t) \), where

\[
v_{\text{tot}} = \sum_{j} v_{ij}
\]  

(8.1)

is the total rate for hopping from site \( i \) to all other sites \( j \).

Because of the exponentially broad distribution of rates, the total
rate will be fairly close to the fastest rate, because the latter will dominate the sum. Thus the probability density $P_{\text{tot}}$ of finding the total rate in an interval around $\ln(v_o/v)$ is fairly close to the probability density of finding the fastest rate in that same interval, which is given exactly by

$$P_1[\ln(v_o/v)] = N[\ln(v_o/v)] \exp\left[-\int_0^{\ln(v_o/v)} N(x')dx'\right]. \quad (8.2)$$

For the fastest rate to be $v$, there must be a channel with that rate, and there must be no faster rate; the probability of the latter is just the exponential of (minus) the expected number of faster channels. This result can be derived by solving the equation

$$P_1[\ln(v_o/v)] = N[\ln(v_o/v)] \left[1 - \int_0^{\ln(v_o/v)} P_1(x)dx\right]. \quad (8.3)$$

As long as the distribution of rates is broad, both the fastest rate distribution and the total rate distribution will be peaked when the argument of the exponential becomes large (see Figure 8.2), that is, at a typical rate, for which the expected number of faster channels is roughly unity:

$$\int_0^{\ln(v_o/v_{\text{typ}})} N[\ln(v_o/v)] \, dv \approx 1 \quad (8.4)$$

This definition of the typical rate, which I call the fastest expectable rate, provides a simple estimate of the total rate that depends only on the distribution of individual rates; it was introduced by Mott (1968) in his derivation of the $T^m$ law for variable-range hopping. Only with the work of Ambegaokar, Halperin and Langer (1971), however, did the nature of the approximation become clear.

We will use the typical rate defined by Eq. (8.4) throughout the rest of the chapter. The states with faster rates are not neglected;
they are simply allowed to contribute only as much as the states with
the typical rate. The true total rate distribution is thus replaced by
the single-rate density $N[\ln(v_o/v)]$, truncated at $v_{typ}$, as shown in Fig.
8.2.

8.3 Zero Temperature

We now consider the case of hopping in an exponential DOS, as in
Eq. (3.17). We begin with the case T=0, for which only downward hops are
allowed. The density of rates available to a carrier with energy $E$ is
then simply related to the total number of deeper sites $N_d(E)$:

$$N_d(E) = \int_E^\infty g(E)dE = N_L e^{E/kT_0}.$$  \hspace{1cm} (8.5)

which are assumed to be randomly distributed in space, so that
$N(R)dR = 4\pi N_dR^2dR$.

According to the naive formula (Eq. (7.1)) for the hopping rate, a
carrier is equally likely to hop to any deeper state, with a rate
$v = v_o\exp(-2\gamma R)$:

$$N[\ln(v_o/v)]d\ln v = N(R) \left| \frac{dR}{d\ln v} \right|$$

$$= \frac{4\pi N_d(E)}{(2\gamma)^3} \ln^2(v_o/v) d\ln v.$$ \hspace{1cm} (8.6)

which is the function used in plotting Fig. 8.2.

The probability that a carrier remains for at least a time $t$ on a
site with energy $E$ is approximately the probability that there is no
site closer than a critical $R_c$:

$$P(E,t) = \exp\left[-\frac{4\pi}{3} N_d(E)R_c^3\right]$$

$$= \exp\left[-\frac{\pi}{6\gamma^2} \ln^3(v_o t) e^{E/kT_0}\right],$$ \hspace{1cm} (8.8)

in which
\[ R_c = (2 \gamma)^{-1} \ln(v_o t) \]  \hspace{1cm} (8.9)

is the radius at which \( vt = 1 \).

This probability changes from one (for deeper states) to exponentially small when the argument becomes greater than unity. The energy at which this occurs is just the hopping-down demarcation energy of equation (7.5), which enters the present calculation as the division between states that have probably released their carriers in time \( t \) and those that have not.

The present discussion is similar to a calculation by Kash (1984), and identical to a recent one by Dunstan and Boulitrop (1984), and probably represents the energy distribution of the carriers reasonably accurately. Unfortunately, it does not allow an accurate calculation of the current, because the carriers near \( E_d \), which are most important for the current, are also those most sensitive to the exact history of the occupation of higher-lying states. We need to know not only the probability that a carrier remains for a time \( t \), but also the total number of carriers that have ever fallen in. This number depends on the exact history of the occupation number, and a reliable calculation would have to be made self-consistent in this regard. Nonetheless, the general features of the energy relaxation must be correct.

### 3.4 Finite Temperatures: Distribution of Rates

We now consider finite temperatures, continuing to restrict our discussion to the simplest case of an exponential DOS and the naive form for the hopping rate. In general, at any temperature there will be a typical rate \( v_{\text{typ}} \) for each initial energy. If a carrier can hop to some state at a rate faster than \( v_{\text{typ}} \), it will do so, but rates slower than \( v_{\text{typ}} \) will not be hopped to, because the carrier will already have gone
somewhere else. The critical radius $R_c$ at which $\nu_{\text{typ}}^t=1$ now depends on energy, and can be determined by inverting the full expression for the hopping rate (Eq. (7.1)):

$$R_c(\Delta E) = \begin{cases} 
0 & \frac{\Delta E}{kT} > \ln(\nu_o/\nu_{\text{typ}}) \\
\frac{[\ln(\nu_o/\nu_{\text{typ}}) - \Delta E/kT]^2}{2\gamma} & 0 < \frac{\Delta E}{kT} < \ln(\nu_o/\nu_{\text{typ}}) \\
\frac{\ln^2(\nu_o/\nu_{\text{typ}})}{2\gamma} & \frac{\Delta E}{kT} < 0
\end{cases} \quad (8.10)$$

where $\Delta E = E_f - E_i$. [In general, the critical radius would depend on both the initial and final energies.] The number of final states per unit energy that can be reached with a rate faster than $\nu_{\text{typ}}$ is just

$$N_f(E_f; E_i, T) = \frac{4\pi}{3} R_c^3(\Delta E, T) g(E_f). \quad (8.11)$$

In determining the distribution of various quantities, we will include only rates that are at least as fast as the typical rate $\nu_{\text{typ}}$. It is useful to define the dimensionless parameter $\gamma$ for this cutoff:

$$\gamma = \frac{T}{T_o} \ln(\nu_o/\nu_{\text{typ}}) \quad (8.12)$$

and to define a dimensionless energy

$$\varepsilon = \frac{E - E_o}{kT_o} \quad (8.13)$$

in which

$$E_o = -kT_o \ln \left[ \frac{\pi N_L}{6\gamma^3} \right] \quad (8.14)$$

is the energy at which the wavefunctions just overlap. As mentioned in the previous chapter, we expect that this energy should not be far from
the mobility edge.

Using these definitions, the density of fast states per unit $\varepsilon_f$ is

$$N_\uparrow(\varepsilon_f) = e^{\varepsilon_1 - \varepsilon_t} \varepsilon \Delta \varepsilon \times \begin{cases} 0 & \Delta \varepsilon > y \\ (y - \Delta \varepsilon)^3 & 0 < \Delta \varepsilon < y \\ y^3 & \Delta \varepsilon < 0 \end{cases}$$

(8.15)

where $\Delta \varepsilon = \varepsilon_f - \varepsilon_i = \Delta E/kT_0$. $\varepsilon_t = E_t/kT_0$ is the dimensionless form of the transport energy

$$E_t = E_0 - 3kT_0 \ln \left( \frac{T_0}{T} \right)$$

(8.16)

Note that the initial energy enters only in the form $\varepsilon_1 - \varepsilon_t$, so $E_t$ serves as a reference energy for this simple expression.

**8.5 Determination of Typical Rate**

Using Eq. (8.16) we can compute various properties of the system as weighted averages over this distribution of final states. The most important quantity determined in this way is the typical rate for each initial energy, since this determines which final states are accessible and thereby all other quantities. The typical rate is determined self-consistently, by setting the total number of final states at all final energies equal to unity, as in Eq. (8.4).

$$1 = \int_{-\infty}^{\infty} N_\uparrow(\varepsilon_f) d\varepsilon_f$$

$$= e^{\varepsilon_1 - \varepsilon_t} \left[ \int_{-\infty}^{0} y^3 e^{\Delta \varepsilon} d\Delta \varepsilon + \int_{0}^{y} (y - \Delta \varepsilon)^3 e^{\Delta \varepsilon} d\Delta \varepsilon \right].$$

(8.17)

In this equation, the first term corresponds to hopping down in energy, the second to hopping up. The hopping-up term is the rate limiting step in the variable-range hopping (VRH) problem, in which the saturation of states below the Fermi level prevents carriers from con-
tinually hopping down. In the classic VRH problem (Mott, 1966) the DOS is taken to be uniform: \( kT = \infty \) and \( N_L/kT = g_o \). If we integrate and invert the first term of Eq. (8.17) in this limit (or simply repeat the integration of Eq. (8.12) for a uniform DOS), we obtain

\[

\nu_{typ} = \nu_0 \exp\left[-(T_1/T)^{\frac{1}{3}}\right].
\]

(8.18)

in which

\[

T_1 = \frac{24\gamma^3}{\pi g_0 k}.
\]

(8.19)

This is the famous "\( T^{\frac{1}{3}} \)" law derived by Mott; it is correctly obtained only when one considers the total number of rates and not the average rate. We set the expected number of rates equal to unity, which appears appropriate for energy relaxation since any hop will allow the carrier to change its energy. If we had used a larger number (~2) appropriate to percolation, the agreement of the numerical value \( 24/\pi \) in the expression for \( T_1 \) would agree almost exactly with the value of 16 obtained by Ambegaokar, Halperin and Langer (1971).

We now return to the case of exponential band tails. In our discussion we will encounter several integrals of the form

\[

I_n = \int_0^\infty d\Delta \varepsilon \, e^{\Delta \varepsilon} \left[y - \Delta \varepsilon\right]^n
\]

(8.20)

\[

= n! \left[ e^y - \sum_{r=0}^{n} \frac{n!}{(n-r)!} y^{n-r} \right].
\]

Using the expression for \( I_3 \) in Eq. (8.17), we find

\[

1 = e^{\varepsilon_1 - \varepsilon_0} \left[y - (1 + y + \frac{1}{2}y^2)\right].
\]

(8.21)

Although this equation cannot be analytically solved for \( y \), it can be solved for \( \varepsilon_1 \), and numerically inverted. Since \( y \) is directly related to
\( v_{\text{typ}} \), we thus have a simple prescription for determining the typical rate for each initial energy. The procedure used could easily be generalized to other densities of states and more sophisticated forms for the hopping rate; for the remainder of this chapter, however, we will discuss only the simplest case.

8.6 Energy Relaxation

In our discussions of dispersive transport, a central quantity was the demarcation energy \( E_d \) at each time \( t \). The preceding discussion allows us to write an analytic expression for \( E_d \), for all times and temperatures, simply by setting \( y \) equal to

\[
y = \frac{T}{T_0} \ln(v_0 t) \tag{8.22}
\]

From Eq. (8.21), we find directly

\[
E_d(t,T) = E_t(T) - \ln \left[ e^y - (1 + y + \frac{1}{2} y^2) \right] \tag{8.23}
\]

We can obtain the hopping-down and hopping-up limits directly from this expression. For \( y(t) \gg 1 \) (long times and high temperatures), the exponential \( e^y \) dominates the expression in braces and we obtain

\[
E_d(t,T) = E_t(T) - kT \ln(t/\tau_s), \tag{8.24}
\]

where, as before,

\[
\tau_s = v_0^{-1} \exp(3T_0/T) \tag{8.25}
\]

as before.

Eq. (8.24) is written in a form that emphasizes the new microscopics driving the multiple-trapping process when hopping is included. It is worth emphasizing that in spite of the change in \( E_t \) and \( \tau_s \) with temperature, the the energy scale is almost unaffected:
Figure 8.3. Demarcation Energy. The time dependence of $E_d$, according to Eq. (8.21) and (8.22), for various temperatures. Also shown are the limiting forms for hopping down and hopping up and the segregation times $\tau_s(T)$ (arrows).

$$E_d(t,T) = E_0 - 3kT_0 \ln(3/e) - kT \ln(\nu_0 t).$$ \hfill (8.26)

For $y \ll 1$ (short times and low temperatures), the algebraic expression cancels terms in the Taylor expansion of the exponential out to third order in $y$. We thus obtain the temperature independent, hopping-down form:

$$E_d(t) = E_t(T) - \frac{1}{6} y^3 kT_0$$ \hfill (8.27)
\[ \mathcal{E}_t = E_0 - kT_0 \ln\left[ \ln(\nu_0 t) \right]. \] (8.28)

Fig. 8.3 shows the calculated form of \( E_d \) for all times at various temperatures, along with the two limiting forms. (Note that the horizontal scale is the natural log of time.) The various segregation times \( \tau_s(T) \) are indicated by arrows. Note that the MT description is valid soon after \( \tau_s \), but only at very short times is the pure hopping-down form a reasonable description.

### 8.7 Energy Distribution

By using the values of \( \nu_{\text{typ}} \) from Eq. (8.21) in Eq. (8.11), we obtain the final state distribution \( N_x \) for various initial energies; this is illustrated in Fig. 8.4. As argued heuristically in the previous chapter, the character of this distribution changes markedly with the initial energy: carriers starting in shallow states carriers hop primarily to lower energies, while carriers in deeps states hop up to the transport energy.

In Fig. 8.4 the initial energies used correspond to \( E_d(t) \), evaluated at \( t=10^n \nu_0^{-1} \), with \( n=0,1,2,\ldots \). Thus, in addition to illustrating the distribution of final energies for each initial energy, the figure shows the evolution from hopping-down relaxation to hopping up relaxation with time. Clearly, once the hopping-up limit is reached the distribution of final states is independent of initial energy. Also evident is the much longer time needed to reach the hopping-up limit at lower temperatures.

Note that there is a significant spread of final energies (several \( kT_0 \)) in all cases; it is clearly an oversimplification to say that the carriers hop to \( E_t \), for example. This warning applies to all "energies" calculated here: they are really just representative energies in
Figure 8.4. Final State Distribution. The density of final states \( N_f \) as a function of final energy for (a) \( T/T_o=0.5 \) (b) \( T/T_o=0.2 \). In each case a set of curves is shown, corresponding to initial energies (arrows) which are the demarcation energies at successive decades of time.

distributions spread over a range of order \( kT_o \). This is a result of their sensitivity to statistical variations in near-neighbor distances; these variations become unimportant only when the DOS changes by a large factor. In spite of its approximate nature, \( E_t \) is a useful indicator of the trends with temperature, and provides a simple way of visualizing the difference between the two types of carrier redistribution.
8.8 Hopping-Down Current

States shallower than $E_d$ are in steady state, with equal numbers of carriers entering and leaving. In the hopping-down regime, however, the shallow states will not be in equilibrium with one another, because the transfer of carriers between shallow states (which would bring them into equilibrium) will not be as fast as hopping to deeper states.

The actual current results from a preference for hopping downfield because of the larger number of energetically accessible downfield states. If we consider $T=0$, then a spatially close state will be accessible only if it is lower in energy than the initial site. Fig. 8.5 shows the origin of the current (see also Section 9.2): the electric field $F$ lowers the energy of downfield sites relative to the initial site, and thus the expected number of downfield sites is larger by a fraction of order $eF R_{typ}/kT_0$ than the number of upfield sites. $R_{typ}$, the typical hopping distance for the initial energy in question, which grows only logarithmically with the time.

Some small fraction of the centers will thus have a previously unavailable center made available in the downfield direction, and will thus execute that hop rather than a slower one. The resulting charge displacement will be proportional to $R_{typ}^2$. Since the number of carriers moving is proportional to $1/t$, the overall current will resemble a power law with slope near $-1$, with logarithmic corrections. Since the redistribution process is not driven by thermal activation, it is expected that the magnitude of this current will be only weakly dependent on temperature.

Clearly this derivation is only approximate, and the logarithmic corrections could potentially change the slope significantly from $t^{-1}$. A
more reliable calculation of the expected current has been performed by M. Grünwald (unpublished) along the lines of Grünwald et al. 1984; the results are in substantial agreement as to the slope, as shown in Fig. 8.6.

8.9 Quasi-equilibrium current

We now turn to the long-time regime, and discuss the magnitude of the current carried by the quasi-equilibrium distribution of carriers.
near the transport energy. In the previous chapter we estimated that the mobility, referred to $E_t$, should be proportional to $\tau_s^{-1}$. However, the energy distributions involved are several $kT_o$ wide, so it is unclear how good this estimate ought to be, since the Boltzmann factor is changing dramatically over the distribution. In this section we will treat the whole distribution rather than just presuming that all the action is at
\( E_t \); this provides important physical insight, as we shall see.

Our discussion applies to both transient and dc conduction, since in both cases the carriers near the transport energy constitute a thermal equilibrium bath. We would like to know the current carried by hopping in this nondegenerate electron distribution. In the absence of a field the occupancy of all shallow states would be given by a Boltzmann factor, \( \exp(-E/kT) \), times the occupation number at the mobility edge \((E=0)\). Naively, one would think that the current per unit energy would be simply the number of carriers per unit energy \( f(E)g(E) \) weighted by the distribution of hops at that energy.

In fact, when we apply a field the states will attain a steady-state rather than an equilibrium occupancy (see Section 9.5). States with an anomalously fast downfield hop will become depleted. Fortunately, as long as we use typical rather than average rates to compute the current, the use of the average occupation number should be reasonably accurate, since the anomalous rates are not included. An alternative procedure, based on direct self-consistent evaluation of the steady state occupancies (the Extended Pair Approximation) has been discussed by Butcher et al. (1983).

The hopping current results from a preference for hops in the downfield direction because of the lower electrostatic potential energy. This energy difference and the charge displacement are both proportional to the hopping distance, so carriers at a given energy will carry a current roughly proportional to the product of their typical hopping rate \( v_{typ} \) and the square of the typical distance they hop, \( R_{typ} \). This result is also obtained by postulating a diffusion coefficient \( D \) for each energy and calculating the mobility using the Einstein relation.
\[ D/kT = \mu/e. \]

The hopping distance can be estimated as follows. Carriers shallower than \( E_t \) will tend to hop the distance necessary to find another state nearby in energy; this distance grows as the inverse cube root of the local DOS, so \( R_{\text{typ}} \sim N_L^{-2/3} \exp(2E/3kT_0) \). Carriers deeper than \( E_t \), however, will tend to hop only the distance to find a center at \( E_t \), which is roughly \( N_L^{-2/3} \exp(2E_t/3kT_0) \), independent of the initial energy.

Figure 8.7. Mean-Square Hopping Distance, from Eq. (8.29)

These heuristic considerations are supported by a direct calculation of the mean-square hopping distance, using the formulas derived
earlier in the chapter. If at a given final energy carriers hop with equal probability anywhere inside a sphere of radius $R_C(E_F)$, their mean-square displacement is simply $\frac{3R_C^2}{5}$. $R_C$ can be related to the typical rate using Eq. (8.21), and the weighted average over the various final energies can be computed using the distribution $N_\gamma$ of Eq. (8.11). The result is

$$\langle R^2 \rangle = \frac{1}{(2\gamma)^2} \left( \frac{\gamma T}{T} \right)^2 e^{\frac{\epsilon_1 - \epsilon_t}{10}} I_4(y).$$

(8.29)

![Diagram showing effective mobility](image)

**Figure 8.8.** Effective Mobility, from Eq. (8.31)

This is illustrated in Fig. 8.7. The display of different tem-
peratures is actually redundant, since Eq. (8.29) shows that \((T_o/T)^2<R^2>\) depends only on \(\varepsilon_1-\varepsilon_t\).

A similar calculation for the first moment of the distance distribution yields

\[
\langle|R|\rangle = \frac{1}{2T} \left(\frac{T_o}{T}\right) \frac{\varepsilon_1}{5} I_3(y). \tag{8.30}
\]

The effective mobility of carriers at an energy \(E\), is estimated to be

\[
\mu(E) \approx \frac{e}{kT} D(E) \tag{8.31}
\]

\[
= \frac{e}{5kT} v_{typ}(E) <R^2(E)>.
\]

This is shown in Fig. 8.8. Note that, for all temperatures, the mobility is peaked at around \(E_0\). The mobility falls off at higher energies because the hopping distance is getting exponentially shorter, while at lower energies the hopping rate is dropping. The mobility near \(E_0\) is nearly independent of temperature because it is dominated by the hopping-down process.

Although the mobility is largest for \(E-E_0\), most of the current comes from carriers originating in deeper states. The product of the effective mobility with the number of carriers \(f(E)g(E)\) gives an indication of the energies at which the current is carried, and is plotted in Fig. 8.9 for two temperatures. We can see that the current is primarily carried around \(E_t\), justifying the name transport energy, although the distribution is several \(kT_o\) wide. There may be some systematic shifts relative to \(E_t\); these are discussed in the next chapter.

The total conductivity is just the integral of \(v_{typ}(E)f(E)g(E)<R^2(E)>\) at various temperatures, and this numerically determined quantity is plotted in Fig. 8.10. It is also possible to
Figure 8.9. Differential Conductivity (current per unit energy).

express this integral analytically, since all the integrands are known functions of \( y \) (although not of energy). We have normalized the integral to the occupation number at \( E_o \), \( f(E_o) = \exp(-(E_o-E_F)/kT) \), in plotting Fig. 8.10. Thus the figure represents the reduction in the temperature dependence resulting from the inclusion of hopping, as compared with band transport at \( E_o \). The dc conductivity is discussed in more detail in the next chapter.
Figure 8.10. Total Conductivity at Different Temperatures.
The quantity shown is normalized to the occupation of states at $E_0$, to show more clearly the changes resulting from inclusion of hopping.

8.10 Inclusion of Extended States

Up to this point we have assumed that the formulas describing hopping rates apply for all energies, even $E>0$. That is, we have not included extended states. There are several reasons for this, one of which is a desire to avoid introducing extra parameters into our discussion. More importantly, our results indicate that for long times the transport in localized states is similar to that expected for extended
state transport. The fact that no extended states were present in the model serves to emphasize this interesting result.

The widely accepted description of the mobility edge as a continuous transition (Abrahams, Anderson, Licciardello, and Ramakrishnan, 1979) suggests that the important quantities for describing transitions to extended states near the edge should be similar to the analogous quantities for transitions to localized states near the edge. We have already seen a similar phenomenon in describing the optical excitation in Chapter 5. Since the maximization of the hopping rate occurred for a final energy \( E_t \) significantly deeper than the mobility edge, the exact behavior of parameters at the edge should not make much difference.

The continuous description is not universally accepted, however. Significantly, Mott, whose intuition has provided much of the basis for our understanding of amorphous semiconductors, does not consider the case proved (Mott and Kaveh, 1983). It is therefore worth examining the case in which the mobility rises dramatically at the mobility edge \( E_c \).

If the mobility rises more rapidly than \( f(E)g(E) \) falls (\( \propto \exp[-E/(kT_0-1/kT)] \)), then the current will have two maxima, one near \( E_t \) associated with hopping, and another near \( E_c \). For times longer than \( \tau_s \), the two sets of carriers will be in equilibrium, and the two channels will act in parallel. The outcome of this competition depends on the difference in the conductivity and on the energy difference between the mobility edge and the transport energy.

We thus return to the traditional picture, in which extended state conduction competes with hopping in band tails. The discussion given by Mott and Davis (1979) should suffice for this case as well, provided one uses \( E_c - E_t \) in place of their tail width. It should be noted, however,
that since $E_t$ is generally several $kT$ below $E_0$, the mobility must rise rather rapidly in order that extended states dominate the conduction. They may play a role in other properties, however.

8.11 Relationship to Isoenergetic Hopping

We have presented the discussion of room temperature dispersive transport totally in terms of the multiple-trapping model, since this model is now widely accepted. For many years, however, the reigning model of dispersive transport in $\alpha$-$\text{As}_2\text{Se}_3$ involved conduction by hopping among a set of isoenergetic traps. The dispersion was viewed as arising from the distribution of hopping distances rather than the distribution of hopping energies. Clearly such a system should exhibit no temperature dependence, and is thus inconsistent with experimental data. The MT model, which agrees with the data, is also much simpler. There has been considerable controversy, however, over the degree of dispersion that is possible for an isoenergetic system. The present model sheds some light on this issue, although no more than was already present in the work of Pollak (1977).

Schmidlin (1980) was the first to divide localized states into transport states and traps on the basis of their kinetics. For the isoenergetic system, he showed that the transport states far outnumber the traps. Since for an isoenergetic system all states have the same occupancy in equilibrium, the traps play no significant role in the transport. Schmidlin therefore concluded that there should be no dispersion in a system of isoenergetic traps. He criticized the work of Scher and Montroll on the grounds that, in that work, states that have slow hopping rates are supplied with carriers just as quickly as fast states are, whereas microscopic reversibility requires that entering and
leaving rates be correlated. Only by having different energies can the two rates be different.

The insight provided by the present model is the importance of the segregation time in dividing two regimes of behavior. For the isoenergetic system this time is \( \tau_s - v_o^{-1}\exp[-2\gamma N^{-1/3}] \), where \( N \) is the density of traps. Although Schmidlin recognized the existence of this characteristic time (which he called \( \tau_c \)) for the transport states, he failed to realize that measurements performed on a faster time scale would be sensitive to the internal redistributions in the transport states.

For a time of order \( \tau_s \) after the application of a field, the occupation of the transport states will be changing towards their steady-state values (which are not their equilibrium values). The current will be larger until steady state is achieved; the situation is exactly the same as in ac conductivity (see Section 9.5). Indeed, for an isoenergetic hopping system the transient response is just the Fourier transform of the ac conductivity. (In an energetically distributed system, the energy relaxation is irreversible.)

Schmidlin's criticism of the Scher and Montroll treatment does appear to be valid, however. The differences in occupancy of different states becomes significant at times of order \( \tau_s \), after which the formalism becomes incorrect (Scher, private communication, has acknowledged the weaknesses in his formulation at long times.) Thus it appears that for times longer than \( \tau_s \) the division of states into transport states and traps is valid, and no further dispersion is expected. this is essentially the description given by Pollak (1977).
8.12 Microscopic Rates

We now discuss the naive hopping rate formula in more detail. The rate of hopping between two localized states depends on the detailed properties of the states in question and their coupling to phonons; the situation is especially complicated if the states differ in energy by many phonon energies. Fortunately, however, because of the broad distribution of rates resulting from statistical fluctuations in distance, we need not know the true numerical magnitude of any particular rate with great accuracy.

The form for the rate in Eq. (7.1) has been widely used in the literature; its main virtue is simplicity, since it ignores several important effects. Although this formula is sometimes attributed to Miller and Abrahams (1960), even those authors included algebraic corrections to the dependence on $R_{ij}$ (Miller and Abrahams, 1960) as a result of the detailed nature of the wave functions.

In application to band tails, Eq. (7.1) is sometimes criticized because it ignores the divergence of the wavefunction decay length near the mobility edge. A crude estimate of this divergence is $(E_{c}-E)^{-\frac{1}{2}}$, which is just the decay length of a localized state in a uniform potential background. More sophisticated calculations give an exponent of 0.6 (Anderson, 1972), or 2/3, instead of 1/2. The particular form does not much matter, however: these divergences can be important only in the immediate vicinity of the mobility edge, since the calculations assume that the DOS is constant. Even $kT_{0}$ below the mobility edge the DOS is substantially smaller and the divergence will be unimportant. Even if the decay length continued to diverge as $(E_{c}-E)^{-s}$, the logarithmic derivative with respect to energy, $s/(E_{c}-E)$ is small compared to the
logarithmic derivative of the hopping distance $1/3kT_0$. Because the hopping distance varies exponentially with energy, this is by far the more important factor for even moderately deep states.

The most severe approximation involved in the naive formula is the simple Boltzmann dependence on energy. The multiphonon nature of the transition suggests the possibility of a more complicated dependence on the energies of the states (see, for example, Stoneham, 1981), as we have already discussed in Section 4.1. In that section we argued that, in the MT problem, the effect of these corrections to the naive rate would be to modify the effective DOS and energy in which the transport was occurring. It may be that such a procedure would be appropriate here as well, but here we have two trap energies to consider. Clearly, however, in the hopping-up (MT) regime, the hopping rates enter only as an average over the distribution of final states. In this case the considerations of Section 4.1 can be applied without modification.

A final comment is in order considering more basic assumptions underlying the naive hopping formula. This formula is obviously rooted in a view of amorphous semiconductors as basically homogeneous, with randomly distributed localized sites governing the properties. Such a viewpoint has been widely successful in describing even such grossly inhomogeneous materials as a-Si:H. Nonetheless, its success in one regime does not guarantee its correctness. It is always possible that the role of long-range potential fluctuations are important; these would cause correlations between the various hopping rates (see Pollak, 1974; Overhof, 1984). Alternatively, the intermediate-range order in a-As$_2$Se$_3$ may cause strongly anisotropic hopping rates; note that the numerous factors of three present in the formulas arise directly from the dimen-
sionality, which could be manifested differently in a layered material. Together with our naive treatment of the multiphonon effects, these comments should serve as a reminder that there is much to be learned about these systems.
CHAPTER 9

IMPLICATIONS OF HOPPING MODEL FOR DC TRANSPORT

In order to explain the low-temperature anomalies in the transient measurements, we were led to include hopping between band-tail states into our description. We found that conduction can occur at a transport energy within the band tail; this process has many properties that have traditionally been associated with extended state conduction. The fundamental transport is by hopping, however, and this has interesting implications for many aspects of transport in disordered semiconductors. In this chapter we discuss some of these implications.

9.1 dc Conductivity

In the last chapter we discussed the calculation of the dc conductivity using the full distribution of hopping rates. Here we return to a simpler discussion that contains much of the important physics.

The dc conductivity $\sigma$ is given by $\sigma = \int f(E)g(E)\mu(E)dE$, in which $f(E)$ is the Fermi function, and $\mu(E)$ is the mobility of carriers at $E_t$. Since the current is sharply peaked near $E_t$ (see Fig. 8.9), we can estimate the temperature dependence of $\sigma$ by

$$\sigma(T) \approx f(E_t)g(E_t)\mu(E_t). \quad (9.1)$$

The mobility should be proportional the hopping rate $\tau_s^{-1}$ at $E_t$,

$$\mu(E_t) \propto \tau_s^{-1} = \nu_0 \exp(-3T_0/T) \quad (9.2)$$

and to the square of the typical hopping distance at $E_t$,

$$R_{typ}^2 = [g(E_t)kT_0]^{-2/3} \quad (9.3)$$

$$\propto (T_0/T)^2 .$$
In addition, there is a factor of $e/kT$ in converting diffusion coefficient to mobility.

Using this expression for the mobility, we find

$$\sigma(T) = (T_o/T)^3 \exp(-(E_t-E_F)/kT) \exp(E_t/kT_o) \exp(-3kT_o/kT). \quad (9.4)$$

The conductivity activation energy $E_o$ is defined to be the derivative of the logarithm of $1/\sigma$ with respect to $1/kT$, so with a little manipulation we find

$$E_o = (E_t - E_F) \quad (9.5)$$

$$= (E_o - E_F) - 3kT_o \ln(T_o/T) \quad (9.6)$$

The intrinsic temperature dependence of the transport process affects the interpretation of all experimental parameters derived from transport measurements. Fortunately, $E_o$ is one of the least affected, and is equal to the true energy for transport, $E_t-E_F$, according to these estimates. Presumably the energies differ by no more than a few $kT$ in a more accurate calculation. The small difference results from a cancellation between the temperature dependence of the transport energy and that of the mobility. The difference between $E_t$ and the mobility edge may be substantial, however. In addition, note that the energy scale for transient measurements is referred to roughly $E_o$ (Eq. 8.26) rather than $E_t$. This means that the energy scales set by the conductivity and the transport measurements will differ by a few tenths of an eV. This is small enough to explain why this effect has not been obvious to experimenters; nonetheless, the change can be important. Fig. 9.1 illustrates the relationship between the various energy scales, according to the simple approximations.

Since the temperature dependence of $E_t$ is nonlinear, one also
Figure 9.1. Relationship Between Energy Scales. Transient measurements measure energies $kT\ln(v_0t)$ below $E_o$, according to Eq. 8.26; the activation energy of the dc conductivity is closer to $E_t-E_F$, which differs from $E_o-E_F$ by roughly $3kT_o\ln(T_o/T)$.

expects some curvature in semilog plots of conductivity versus $1/T$. Although there are numerous reports of such curvature in the literature (Callaerts, Nagels and Denayer, 1972; Sharp, Marshall, and Fortuna, 1981), some of these reports no doubt reflect the extraordinary difficulty of avoiding experimental artifacts in measuring very high resis-
tances; some of the most careful work (Seager and Quinn, 1975) indicates purely activated behavior. Since the expected changes in activation energy (of order \( kT_0 \), for \( T \) near \( T_0 \)) are small compared with the activation energy (\( \sim 1eV \)), however, it is conceivable that the slight curvature may have been overlooked.

Another open question concerns the magnitude of the prefactor of the conductivity, which of course requires a knowledge of other parameters besides \( kT_0 \), for example the hopping rate prefactor \( v_0 \) and the wavefunction size \( \gamma^{-1} \). The conductivity prefactor will be a (small) multiplicative factor times

\[
\sigma_0 = \frac{v_0 Ye^2}{kT_0}
\]

(9.7)

Using \( v_0 = 2 \times 10^{12} s^{-1} \), \( \gamma^{-1} = 5 \AA \), and \( kT_0 = 50 \text{meV} \), the prefactor is of order \( 100(\Omega \cdot \text{cm})^{-1} \). Hopping models tend to give prefactors two or three orders of magnitude smaller than those experimentally observed. The polaron model predicts a somewhat larger value because of the shorter distance (Seager and Quinn, 1975), but generally still too small.

In evaluating the magnitude of the prefactor, one must take into account the possibility of a temperature dependence of the activation energy \( E_a \) (Davis and Mott, 1970). If the variation of \( E_g \) is linear over some temperature range, \( E_g(T) = E_g(0) + YT \),

\[
\sigma(T) = \sigma_0 e^{-E_g(T)/kT}
\]

(9.8)

\[
= [\sigma_0 e^{-\gamma/k}] e^{-E_g(0)/kT}
\]

We have already included the temperature dependence of the band-tail transport in our equation for \( E_g \); we have not, however, include the possibility that the bands themselves are moving relative to the Fermi
level. One really has no idea of the temperature dependence of $E_0 - E_F$, so the magnitude of the prefactor is a difficult quantity to interpret. Often $\gamma$ is estimated from the temperature dependence of the optical absorption, which gives an estimate of around 10 for $e^{-\gamma/k}$ (Davis). This still leaves the magnitudes somewhat small, although the highest estimates for polaronic effects in the context of the temperature dependence of the hopping rate arising from the multiphonon nature of the transitions, may also have some relevance.

9.2 Field Dependence of Conductivity

The conductivity of many amorphous semiconductors, including a-As$_2$Se$_3$ (de Wit and Crevecoeur, 1972) depends exponentially on electric field strength $F$. Such an effect could in principle arise from a variety of effects, for example, diode-like contact resistances. Comparison of well-characterized samples of different thicknesses, however, shows that the field dependence is a true bulk effect. In addition, Marshall and Miller (1973) examined several models involving, for example, field modulation of trapping rates in a trap-limited mobility and concluded that they could not explain the data. The authors were forced to conclude that there was an intrinsic field dependence to the microscopic mobility, even though they had no microscopic model for such an effect.

Such a field dependence is in fact expected from the model of hopping in exponential band tails. Recall that the mechanism driving the current is the larger number of accessible downfield states. This fractional asymmetry is the number of new states made available at the hopping distance $R_{typ}$ relative to the total number, so
Figure 9.2. Mechanism for Exponential Field Dependence. The field makes more states energetically accessible on the downfield side, fewer on the upfield side.

\[
i(F) \propto \frac{\int_{E-eFR_{typ}}^{E+eFR_{typ}} g(E') dE'}{\int_{-\infty}^{E} g(E') dE'}
\]

\[
\propto \exp \left| \frac{eFR_{typ}}{kT_0} \right|
\]  

(9.9)  

(9.10)

as shown in Fig. 9.2. The exponential field dependence is thus a natural manifestation of hopping in an exponential DOS. The dependence of the transport energy, and thus the hopping distance, in temperature results
in a stronger field dependence at low temperatures, as observed experimentally. The hopping picture thus provides a simple explanation for an ubiquitous effect.

9.3 Thermopower

A longstanding puzzles in the dc conduction of amorphous semiconductors has been a low value for the slope of the thermopower plotted against reciprocal temperature, $E_s$, as compared with the conductivity activation energy, $E_0$. This discrepancy has been regarded (Emin, Seager and Quinn, 1972) as strong evidence that the carriers are small polarons. The difference $E_s - E_0$ then represents the activation energy for hopping of polarons.

Other explanations of the discrepancy (Callaerts, Nagels and Denayer, 1975; Mott and Davis, 1979) involve a change from band conduction to hopping in band tails as temperature is lowered. Hopping in the assumed algebraic density of states (DOS) gives conduction essentially at the bottom of the band (Grant and Davis, 1974; Weiser, Grant, and Moustakas, 1974); thus the transition between processes is fairly discrete. In this model it appears to be coincidental that the change in conduction mechanism should occur in the experimental temperature range. Although these models adequately describe the data, several free parameters are required to fit the data.

In contrast, the model of hopping in exponential band tails developed here explains the thermopower/conductivity discrepancy semi-quantitatively with no additional adjustable parameters. This model involves a continuous change of the mechanism with temperature, as the transport energy moves through the band tail.
The thermopower $S$ measures the mean energy above the Fermi level at which current is carried divided by the temperature (Fritzsche, 1971). We estimate this to be $(E_t - E_F)/T$, where $E_F$ is the Fermi energy. The thermopower slope $E_s$ is defined to be the derivative of the thermopower with respect to $1/T$, so, taking the derivative of $E_t$ from Eq. (7.2),

$$E_s = (E_t - E_F) - 3kT_0$$

(9.11)

$$= (E_o - E_F) - 3kT_0(1+\ln(T_0/T))$$

(9.12)

The difference $E_o - E_s$ is $3kT_0$, or roughly 0.2eV for As$_2$Se$_3$ at 400K. This is to be compared with the value of -0.3eV obtained experimentally by Seager and Quinn (1975). Thus the simple model predicts effects of the right order of magnitude to account for the discrepancy, without polaronic effects. This does not rule out the possibility of an additional contribution because of electron-phonon coupling, however.

To get a better estimate of the discrepancy predicted for exponential band tail model, we can average over the distributions as we did for the conductivity. Specifically,

$$S = \int_{E_F}^{\infty} u(E_1) f(E_1) g(E_1) \left[ \frac{E_m(E_1) - E_F}{kT} \right] dE_1$$

(9.13)

This is the conductivity-weighted average of the quantity $E_m(E) - E_F$, which is just the average amount of energy that moves with the carrier when it hops from $E$ to the various final states. For hops between states with the same energy, this is just their common energy. Unfortunately, for hops between states of different energies, this is a rather slippery quantity.

To examine the question more deeply, we examine first the strongly-coupled case, illustrated in Fig. 9.3., in which the carrier hops between two traps with the same energy and the same spring constant
(curvature). We now need two configurational coordinates, \( Q_1 \) and \( Q_2 \), one for each of the traps. We assume that the curvature is the same for both states, and that the two coordinates do not directly interact. Fig. 9.3 shows the situation in plan view; the suppressed vertical axis is the total energy of the system, which includes contributions from the elastic energy at each site as well as the electronic energy.

If the electron starts at site 1, then the lattice configuration will tend to resemble that in the lower left hand corner. The probability of the system being found in any part of the well is proportional to a Boltzmann factor of the total energy. The electron will not make the transition unless energy is conserved; clearly this occurs most readily at the "pass" between the two wells. A trajectory through the parameter space of the system including this pass is shown on the right. In chemical language \( \bar{Q} \) would be the "reaction coordinate"; we can take \( \bar{Q} = Q_1 - Q_2 \). The pass is called the "transition state," which has aspects of both the initial and final configuration, and is given by \( \bar{Q} = Q^* \). Setting

\[
E_1 = \frac{1}{2} k [Q_1^2 + Q_2^2] = kQ^2
\]

equal to

\[
E_2 = \Delta E + \frac{1}{2} k [(Q_1-\Delta Q)^2 + (Q_2-\Delta Q)^2] = \Delta E + k(\bar{Q}-\Delta Q)^2
\]

along the \( \bar{Q} \) line yields

\[
Q^* = \frac{1}{2} \frac{\Delta E}{k\Delta Q} - \frac{1}{2} \Delta Q. 
\]

When the electron moves from site 1 to site 2, it changes the equilibrium configuration for the atoms at the two wells, since site 1 is now empty and site 2 is now occupied. As a result, some bonds change, for example, from being shorter than the relaxed length to being longer than the new relaxed length. The relevant quantity is the change
Figure 9.4. Hopping between strongly-electron-phonon coupled centers. (a) Equal Energies (b) Different Energies. Left side shows top view of total-energy surface; right side is cross section along diagonal line.

in total energy on, say, site 1, which is just

$$\Delta E_1 = \frac{1}{2}k\bar{Q}^2 + \frac{1}{4}k(\bar{Q} - \Delta Q)^2$$ (9.16)
\[ = \frac{1}{2} \Delta E \]  

(9.17)

Thus in the strong-coupling limit, the energy transferred \( E_m \) is just the average of the two relaxed site energies. This is the expression used by Weiser et al. If this were true, however, one would expect substantial barriers to motion, which would also contribute to the thermopower/conductivity discrepancy as in the polaron case. Since the hopping process itself accounts for most of the discrepancy, these effects cannot be too large.

In addition, several other observations suggest the absence of substantial electron-phonon coupling for the charge carriers, at least in a-As\(_2\)Se\(_3\). For example, the activation energy of the drift mobility (Pfister and Scher, 1978) is simply that predicted by the multiple-trapping model (OKV). An additional activation energy of 0.3eV associated with the microscopic mobility could be observed. In contrast, the hopping model predicts that the multiple-trapping result should be valid in spite of the temperature-dependent mobility. In addition, photo-induced optical absorption measurements exhibit a threshold at \( \hbar \omega = kT \ln(v_0 t) \) as predicted by multiple trapping; in the presence of strong electron-phonon coupling this threshold would be substantially larger.

In the weak-coupling limit, in which the coupling is weak, the energy will be approximately equal to the higher of the two electronic energies, since the energy of the lattice is playing a minimal role. This corresponds to vertical transitions between flat configurational-coordinate curves; the electron then carries all the energy along with it.

In evaluating the thermopower using the distributions of the previ-
Figure 9.4. Relevant energies for various temperatures, including the transport energy, the average final energy of hops, $kT$ times the logarithm of the conductivity, and the weak-coupling thermopower.

In this chapter, I have used the weak-coupling estimate that $E_m$ is the greater of $E_1$ and $E_F$. $E_m$ can then be calculated using a similar approach as was used for the conductivity; a similar approach could also be used for the strong-coupling limit. The results of the calculation are shown in Fig. 9.4, together with other results of the complete calculation.

The general trends in the temperature dependence of the thermopower follow those of the logarithm of the conductivity. Interestingly,
although these transport-related quantities seem to have a similar dependence on temperature, this dependence differs from that of the energy-related quantities $E_t$. This suggests that the description in terms of a single transport energy is oversimplified, but perhaps two would suffice.

9.4 Hall Effect

In crystalline semiconductors the Hall effect provides a useful measure of the carrier density. In amorphous semiconductors it is highly anomalous; the Hall mobility is small, and the sign of the Hall effect is often opposite to that of the thermopower, which is a more reliable measure of the sign of the majority carrier. This phenomenon is one of the strong indications that the transport in amorphous semiconductors is fundamentally different from that in crystals.

The primary difference between the present picture and the mobility edge picture is that the fundamental transport event is hopping, rather than scattering. In this framework, it is more reasonable that the Hall effect should be anomalous.

The Hall effect measures the interference of wavefunctions on paths that cross magnetic flux lines. The successful explanations of the Hall effect sign anomaly invoke closed loops involving odd numbers of sites (Friedman, 1971; Grünwald, 1981). The sign of the Hall effect also depends on the relative phase at the different sites of the wave function corresponding to the band extremum: if the electron wave function is in phase, then the magnetic field will cause destructive interference, and the normal Hall effect will result. Since hopping involves interaction with phonons, it destroys phase coherence, and one expects the Hall effect to be small.
Alternatively, some of the Hall effect may be contributed by carriers moving in extended states. One puzzle is that heavily-doped crystalline semiconductors, which are in many ways similar to amorphous semiconductors, do not exhibit an anomalous sign of the Hall effect. One possible explanation is that for these materials, conduction by extended states is much larger relative to the hopping contribution; the details of the competition, discussed in Section 8.9, may then be relevant.

9.5 ac Conductivity

The segregation time \( t_s \) is the time scale on which the transport states come into internal equilibrium. Given this fundamental importance, we might expect that this time would have other manifestations than a change in the time dependence of the transient photocurrent, once the temperature is low enough that this time is measurably long. The most promising possibility seems to be the ac conductivity. The theory of this effect is complicated, and the experiments are difficult and incomplete. Because of the importance to the model, however, we will give a brief sketch of the ideas involved.

The equations for the occupations of various localized states can be written:

\[
\frac{df_i}{dt} = \sum_j \left\{ \nu_{ij}f_j - \nu_{ji}f_i \right\} . \tag{9.18}
\]

In equilibrium the net rate of hopping between each of the pairs \((i,j)\) is 0, by the principal of detailed balance. Thus \( \nu_{ij}^o f_j^o = \nu_{ji}^o f_i^o \), where the \( o \) superscript represents equilibrium. We can reframe \( f_i \) in terms of the energy of the state \( \varepsilon_i \) and the chemical potential \( \mu(\overline{r}_i) \) at the site:

\[
f_i = \exp\left\{-[\varepsilon_i - \mu(\overline{r}_i)]/kT\right\} . \tag{9.19}
\]
In the presence of an applied electric field \( \vec{F} \), the energies \( E_i \) are altered by the local electrical potential energy:
\[
\Delta E_i = -e\phi(\vec{R}_i) = +e\vec{F} \cdot \vec{R}_i.
\]
This results in an asymmetry in the hopping rates:
\[
v_{ij}f_i = v_{ji}f_j \exp(\vec{F} \cdot \vec{R}_{ij}/kT).
\]

Figure 9.4. Equivalent circuit for hopping (a) hopping microscopics (b) Equivalent circuit (c) frequency response.

In the limit that of small fractional changes in occupancy, Miller and Abrahams (1959) showed that Eq. (9.18) has the simple equivalent
circuit illustrated in Fig. 9.4 there is a simple, physical mapping between the parameters of the equivalent circuit and that of the true system: the applied voltage $V_i$ is just the local electrostatic potential $\phi(R_i)$, the voltage at the node is $(-1/e)$ times the electrochemical potential $\mu_i - e\phi(R_i)$; the charge on the capacitor $C_i$ is $(-e)$ times the occupation of the state $f_i$; the conductance between $i$ and $j$ is $(-e/kT)$ times the equilibrium flux between the states $\nu_{ij}^O = \nu_{ji}^O$, and the capacitance at each site is $(-e)$ times the equilibrium occupancy of the site.

An individual pair of sites thus constitutes a simple network, as shown in Fig. 9.4. The contribution of such a pair to the conductivity is simply proportional to the projection of $R_{ij}$ along the direction of the field (see Appendix D), and to the current flowing from $i$ to $j$. The current flowing is constant and in phase for $G_{ij}^{-1}(C_i + C_j)\omega > 1$, but falls of linearly with frequency and is capacitative for higher frequencies. Note that the RC time constant is essentially the reciprocal of the slower of the two rates $\nu_{ij}$ and $\nu_{ji}$. In terms of the original picture, and in time domain, the current flows continuously until the occupancy of the more occupied state is sufficiently depleted to offset the asymmetry in fluxes created by the field.

This is the origin of the classic electronic description of the ac conductivity in amorphous semiconductors (Pollak and Geballe, 1961). They assumed that each pair of states contributes individually to the conductivity, with an equivalent circuit such as that in Fig. 9.4. The ac conductivity is then the sum over various distances of the various pair responses; an exponential dependence of the rate (and hence the RC time constant) on the distance gives the $\omega^1$ conductivity; the variation
of the number of sites with distance means that more pairs contribute at lower frequencies; the resultant logarithmic corrections and cause the conductivity to go as roughly $\omega^{0.8}$. Other mechanisms can also explain the frequency dependence of the conductivity (see, for example, Elliot 1979), and in my opinion are at least as plausible; nonetheless, the pair picture is a popular one.

Figure 9.5. Equivalent circuit for a large number of hopping sites.
This description fails to yield the appropriate dc limit to the conductivity, however. This results from the fact that the treatment of each pair as isolated cannot result in conduction of current across the sample, which is required for dc conductivity. At some time range the interconnection between pairs must be as important as the connection within pairs; at this point the pair approximation breaks down.

To address this problem, Butcher and Summerfield (1981; see also Summerfield and Butcher, 1982, 1983) introduced the Extended Pair Approximation (EPA). Each pair is now connected to other states as well, as it must be (see Fig. 9.5). Rather than including the full network, however, the admittances of the connecting network are determined self-consistently. The result is that the conductance rolls over smoothly from the pair-approximation result to a dc current determined by the percolation of the conduction path.

The conductivity thus passes from a frequency-independent regime at low frequencies to a frequency-dependent regime at higher frequencies. The increase with frequency begins when the conductivity contributed by nonpercolating paths, which become conductive at a finite frequency \( \nu_c \), becomes comparable to the conductivity contributed by the percolating path. Since \( \tau_s \) is presumed to be the characteristic hopping time for states in the percolating network (the transport states), we make the following conjecture: frequency \( \omega_{dc} \) is of order \( 1/\tau_s \).

Experimental attempts were made to observe the effects discussed in this section. There is a large frequency-dependent conductivity in the dark; if the present conjecture is correct this cannot come from electrons in band tails, but must arise from some other source such as atomic motion. To isolate the electronic part, one can take the differ-
Figure 8.6. Total Frequency dependence of conductivity. At low frequencies the conductivity comes from a percolating path through the system; for $\omega > \omega_{dc}$ this contribution is exceeded by contributions from a variety of small isolated clusters, primarily pairs.

ence between the conductivity with and without photoconductive illumination. As long as the quasi-Fermi level is deeper than the transport energy (which may not be true at low temperatures) the frequency dependence of the conductivity should be the same for photoconductivity as for dark conductivity. Indeed, our observation, which confirms observations of Abkowitz, Lakatos, and Scher (1974) of frequency independent
photoconductivity at room temperature strongly suggests that the ac con-
ductivity is not an electronic effect, or at least does not involve the
same electrons as the dc conductivity.

At low temperatures the measurement is complicated by the fact that
the photoconductivity takes a long time to reach steady-state. Since
measurements at different frequencies will have different exposure his-
tories, it is hazardous to compare the difference caused by light at
different frequencies. This problem can be overcome by measuring the
phase $\theta$ of the conductivity, which can be related to the frequency
dependence of the magnitude $|\sigma|$ of the complex conductivity using the
Kramers-Kronig relations (Kittel, 1976, p 324 ff.),:

$$\theta(\omega) = -\frac{1}{4\pi} \ln s + \frac{\omega}{s - \omega} \frac{d\ln |\sigma(s)|}{ds} \, ds$$

(9.20)

Since the logarithm is peaked around $s = \omega$, the phase essentially meas-
ures the local frequency dependence of the conductivity.

We therefore expect that the phase of the photoconductive part of
the conductivity should show a phase change at the segregation time; as
the temperature is changed, the frequency dependence should also change
as the segregation time moves. Attempts were made to measure this
effect, but they reached no definitive conclusion by this writing; this
may well prove an interesting subject for future research. The primary
experimental problem encountered was the interference of the pri-
marily capacitive signal from the sample with the much smaller changes
caused by the light; this problem might be solved with a good bridge
circuit. Furthermore, preliminary indications were that the effect was
more gradual than the simple model suggests, less than a 20° phase
shift. This could be reasonable if the transition is spread over several
orders of magnitude (notice that the conductivity saturates again at
high frequencies in Fig. 8.6) so more careful measurements are required. It should be stressed that the observation of a correlation between the temperature dependence of the ac conductivity and of the transient transport would provide definitive proof that the low temperature effects truly constitute a change in the conduction mechanism, as asserted.

9.6 Overview of the Hopping Picture

In summary, we have shown that a simple model of hopping in exponential band tails naturally explains not only transient measurements, but also the anomalous features of dc transport in amorphous semiconductors. The generality of this result, requiring only the existence of more-or-less exponential band tails, explains the ubiquity of anomalous transport in amorphous materials. The model still contains a sharp transport energy, similar to the mobility edge that has provided the impetus for so much recent progress. Nonetheless, the fundamentally temperature-dependent nature of the hopping transport has profound implications for the detailed interpretation of experimental parameters. The simple formulas presented here provide a basis for estimating these effects.

The model explains the circumstances (t >> τs) under which a disordered semiconductor can be treated as having transport states and traps, a common description for room temperature properties. On the other hand, one can easily imagine the circumstances under which hopping is the dominant mechanism of dc transport (low enough temperature so that the transport energy reaches the Fermi energy and becomes infinite); this is the VRH regime.

The hopping mechanism of conduction and the temperature-dependence of the transport energy may well be reflected at all temperatures in the
anomalous behavior of many transport properties. The full importance of these result to the interpretation of experiments is still being explored.
TRANSIENT TRANSPORT AND OPTICAL STUDIES
OF CHALCOGENIDE GLASSES

by

DONALD PAUL MONROE

B.S., Physics, Massachusetts Institute of Technology
(1980)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January, 1985

© Massachusetts Institute of Technology 1984

Signature of Author    Donald P. Monroe
Department of Physics

Certified by    Marc Kastner
Marc Kastner
Thesis Supervisor

Accepted by    George Koster
Chairman, Department Committee

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JAN 25 1985
LIBRARIES

ARCHIVES
CHAPTER 10

FUNDAMENTAL EXCITATION

In this chapter we discuss the primary optical absorption process in amorphous semiconductors. In later chapters we will see dramatic differences in the TPA excited with photons of different energies, so it is important to have some appreciation of the excitations produced at different energies. In addition, this chapter presents the a measurement of the quantum efficiency for creation of photocarriers obtained using transient photocurrent; this technique avoids some of the problems associated with a steady-state measurement. The results are quite intriguing, and their implications for various models will be discussed.

10.1 Overview of Absorption Spectrum

Fig. 10.1 shows the absorption coefficient $\alpha$ of a-As$_2$Se$_3$ on a logarithmic axis, versus photon energy; the spectrum is similar for other amorphous semiconductors, except for differences in the energy scale. It is customary to identify three spectral regions, which we denote A, B and C; these three regions presumably correspond to different physical processes, indicated in Fig. 10.2, as we now discuss. The reader may wish to review Chapter 5 on the microscopic nature of the absorption process.

Region A- Band-to-band absorption. The high absorption coefficient region, $\hbar\omega \lesssim 2.0$eV, involves excitations of electrons from the main part of the valence band to the main part of the conduction band. This conclusion follows from the observation that an absorption coefficient of $10^5\text{cm}^{-1}$ requires the participation of $10^{21}\text{cm}^{-3}$ electrons, if they have a
Figure 10.1 Absorption spectrum of a-As$_2$Se$_3$. For $\hbar \omega < 1.6$eV, this is data for the same sample used for photocurrent and induced absorption measurements.

representative cross section of $10^{16}$cm$^2$. The data can be fitted by the Tauc formula (Kosek and Tauc, 1970),

$$\alpha(\hbar \omega) \propto (\hbar \omega - E_g)^2/\hbar \omega$$  \hspace{1cm} (10.1)

This formula can be rationalized by assuming a constant displacement matrix element (no selection rules) between bands having a
Figure 10.2. Physical processes normally associated with the three spectral regions of absorption indicated in Fig. 10.1: A- Band-to-band excitation. B- Band-to-tail or tail-to-band excitation. C- Defect-to-band or band-to-defect excitation, or direct excitation of internal parabolic DOS as found in crystals. Fitting of the high-photon-energy absorption spectrum to the Tauc form then determines the "optical gap" $E_g$. However, there is really no theoretical justification this form, so the relationship of this gap to, for example, the extrapolated band edges mentioned in Chapter 1 is far from clear. Cody et al., (1982) have concluded that, at least in a-Si:H, a constant momentum matrix element—which leads to a multiplicative rather than a divisive factor of $\hbar \omega$—provides a fit over a broader range. The sensitivity of the resultant $E_g$ demonstrates the need for a better understanding of the physical
processes involved before putting too much weight on the parameters determined.

**Region B—Exponential Absorption Edge: "Urbach" region**

For $1.5\text{eV}<\hbar\omega<2\text{eV}$, the optical absorption depends exponentially on photon energy. This region is normally associated with transitions between band states and tail states. However, since exponential absorption edges are also observed in crystals, where band tails are not expected, the many other possible mechanisms also deserve attention. A detailed discussion is presented in the next section.

**Region C—Low-energy absorption: "Tauc Region"**

For low photon energies the absorption does not continue to decay with the steep exponential, but shows a low-energy tail, with absorption coefficients of order $1\text{cm}^{-1}$. For well-prepared material, this is true absorption, rather than scattering losses from inhomogeneities, as demonstrated by the careful work of Wood and Tauc (1972); for more poorly prepared samples, however, the possibility of losses due to scattering from inhomogeneities should be kept in mind. Wood and Tauc also showed that the magnitude of the low-energy tail was sensitive to sample preparation conditions as well as the presence of impurities. This observation supports the notion that the low-energy tail is associated with defects in the material. If their optical cross section is $10^{16}\text{cm}^2$, then their density is of order $10^{-16}\text{cm}^{-3}$. At least this many defects are expected from other considerations (see Chapter 11). The broad, featureless form of the absorption coefficient is not well understood, however, in part because of the difficulty of making reliable measurements. It may involve excitation from a defect level to a band, or an internal excitation of the defect. The observations of photocurrent discussed below support the former
interpretation, while luminescence data tend to support the latter.

10.2 Models for the Exponential Edge

A more detailed understanding of the nature of the initial excitations resulting from absorption in the exponential region would clarify several aspects of the subsequent behavior of those excitations. Unfortunately, there is a plethora of potential models for exponential edges, and there is no consensus on the appropriate model for amorphous semiconductors. What is clear is that to interpret the exponential edge simply as a DOS effect poses great potential hazards.

The exponential edge is commonly referred to as the "Urbach Edge," since it was first described by Urbach (1953) in observations of crystalline AgBr. Strictly speaking, however, the Urbach rule specifies that the width of the exponential should be proportional to temperature; the exponential tails observed in amorphous semiconductors satisfy only the "spectral Urbach rule" since the widths are only weakly dependent on temperature. This suggests that in amorphous semiconductors the frozen-in disorder is at least as significant as thermal disorder in determining the exponential width. A naive model incorporating this effect has been proposed by Cody et al. (1982).

Models for exponential tails abound, in part because such tails are observed in a wide variety of systems. The most fundamental difference between the various models is whether the broadening leading to the edge is homogeneous (everywhere the same) or inhomogeneous. For amorphous materials, the simplest explanation of the exponential edge involves only the joint density of states (JDOS), with no strong variation of matrix element effects or electron-phonon coupling:

\[
\alpha(\hbar \omega) \propto | \mathbf{f} |^2 \int_{E_F - \hbar \omega}^{E_F} g(E_i) \ g(E_i + \hbar \omega) \ dE_i .
\]  
(10.2)
(Because the tails are narrow, the difference between position and momentum matrix elements is not important.)

It is easy to show that for moderately low energies the JDOS is dominated by transitions between the broader of the two tails and the "shoulder" of the other band (at which the DOS stops varying exponentially), as indicated in Fig. 10.2. In a-Si:H, both the valence- and conduction-band tail widths are known, and the width of the absorption edge is indeed similar to that of the wider band tail. Redfield (1982), however, has pointed out that the absorption exponential is wider than the broader tail (that for holes) and this cannot be explained without including some variation of the matrix elements. These effects must be such that they make low-energy transitions more likely— a surprising result since the models we are about to discuss have larger matrix elements for higher energies.

Such matrix element effects are the standard explanation of the exponential in crystalline semiconductors, where band tails are not generally discussed. The increase of the edge width with temperature suggests that it is a manifestation of the disorder, but there are many mechanisms by which disorder can affect the electronic states, including direct interaction of excitons with phonons (Segall, 1966) or with electrostatic potential fluctuations (Dow and Redfield, 1971, 1972), or incipient self-trapping of excitons (Sumi and Toyozawa, 1971). For an overview, see Mott and Davis (1979, p. 272ff).

The ubiquity of exponential edges suggests the possibility of a single model to explain all the observations, although the real world need not be so simple. The model that comes closest to achieving widespread applicability is due to Dow and Redfield (1971, 1972), and
Figure 10.3. Dow-Redfield model for exponential edge. The horizontal axis in the diagram is the center-of-mass coordinate in the direction of the field. In the presence of an electric field, the excitonic states are mixed with ionized states, making possible transitions with significantly lower energy.

Involves broadening of excitons absorption by internal electric fields, illustrated in Fig. 10.3. It is well known that an excited electron-hole pair have their lowest energy when their motion is strongly correlated, that is, they form an exciton. In the case where the final state is highly correlated, the absorption coefficient is proportional to the square of \( \Psi(0) \), which is the probability amplitude that the electron and hole are at the same place so that \( \bar{r}_e - \bar{r}_h = 0 \).

In zero field, absorption by the lowest excitonic level would give a sharp peak in the absorption. In a strong field, excitations can be
made with substantially lower energy, since absorption can now occur into states in the continuum. The mixing of the continuum with the exciton gives this process a much greater probability than direct excitation into the continuum (the Franz-Keldysh effect), because of the large amplitude of the final state with \( \bar{r}_e = \bar{r}_{h} \). It also changes the exponent of the energy from \( 3/2 \) to \( 1 \) (because the geometry has changed), so purely exponential behavior is expected even for a uniform field (the model can be purely homogeneous). An alternative description is that a low-energy photon can excite a virtual exciton as long as the exciton can be ionized by the field before too much time has passed. In this model the disorder acts only to cause a strong internal electric field. The generality of the model arises from the many mechanisms by which this field can be generated, which of course weakens the case for universality.

The importance of excitonic and phonon effects on the absorption spectrum is suggested by the observations of Sussmann, Searle, and Austin (1981), who saw sharp excitonic and phonon structure in the absorption and electroabsorption spectra of crystalline \( \text{As}_2\text{Se}_3 \). The absorption spectrum exhibits clear step-like features at low temperatures; by room temperature, however, the spectrum has broadened into a featureless, exponential tail. While it is conceivable that the exponential comes from a different source than the structure, the similarity in scale is suggestive.

10.3 Excitation Energy Dependence of Luminescence in Glasses

There is a limit to what one can learn about the nature of the excitations simply by studying the absorption— as we have seen, very different models can produce an exponential edge, for example. Changing
the conditions of the sample, for example the temperature or the applied electric field (electroabsorption) provides more information. To learn still more, one must study the eventual fate of the excitations produced by the absorption; such measurements are inherently affected by the (unknown) dynamics of these excitations, but they provide an important window on the excitation process. In the present section we discuss the photoluminescence (PL). Later sections address the photocurrent (PC). These probes are to a certain extent complementary, since the PC measures electron-hole pairs that have been separated, while the PL measures those that remain correlated throughout their lifetime.

Chalcogenide glasses exhibit a broad PL peak with an energy near half of the optical gap, which can be excited by light with \( h\omega_x \) comparable to the gap. Interestingly, higher energy light is not as effective in exciting luminescence: the quantum efficiency \( \eta_{PL} \) (luminescence photons per absorbed photon) falls at high energies. Street (1974) showed that this fall-off can be described as roughly inversely proportional to the absorption coefficient. This suggests a mechanism in which a constant-absorption-coefficient process resulting in PL competes with the exponential absorption process for excitation photons; the exponential absorption apparently does not give luminescence. This picture suggests that the PL results from direct excitation of luminescence centers, which behave differently from the bulk of the material. Absorption by luminescence centers may also be responsible for the low-energy absorption tail. The necessary Stokes-shift of the PL, from an energy comparable to the bandgap to half of the bandgap, is sufficient to explain most of the width of the PL band as well (see Fig. 10.4).

Alternatively, the apparent inverse proportionality to the absorp-
Figure 10.4. Configurational-coordinate diagram for explaining luminescence. Luminescence occurs from the strongly distorted configuration so that $\hbar \omega_{\text{PL}}$ is substantially less than the bandgap.

tion coefficient could be a coincidence, with the falling $\eta_{\text{PL}}$ resulting from the greater energy of the excitations. This possibility is made more plausible by the observation (in a-$\text{As}_2\text{Se}_3$ but not a-$\text{As}_2\text{S}_3$) that $\eta_{\text{PL}}$ also falls off at low photon energies, which should not happen if only PL center absorption were involved. Although this effect was not reported by Street, it was noted by Hudgens and Kastner (1978); I have confirmed their result using transient measurements.

One mechanism for decreased quantum efficiency would be that a large internal energy makes the excitons more prone to ionization and subsequent non-radiative recombination. Another possibility is that
excitation with different energies in the absorption edge creates excitons in fundamentally different local environments (inhomogeneous broadening). The environments corresponding to higher-energy absorption must then be less conducive to luminescence.

Whatever the details of the model, it appears that excitation in the exponential tail (at least in the low-energy part) creates highly correlated (excitonic) electron-hole pairs, with an efficiency close to 100% (the low temperature $\eta_{PL}$). Such observations seem inconsistent with the Dow-Redfield model, since in that picture the low-energy excitations owe their very existence to their eventual ionization by the field. Excitation at higher energies does not produce luminescence, but whether this is because there are fewer excitons created per absorbed photon or because they ionize is not clear.

It is interesting that the same photon energies that efficiently create PL also induce metastable induced absorption centers discussed in the next chapter. This suggests a relationship between the distortion associated with the Stokes shift of the PL and the rearrangement causing the metastability.

10.4 PL in Crystalline Chalcogenides: Self-Trapped Excitons.

Important insights into the excitation process have been gained in the past few years by studies of the luminescence in crystalline chalcogenide semiconductors. This work is still in progress, and it is far beyond the scope of this work to discuss it in detail. Nonetheless, it is useful to examine the extent to which a coherent picture can be generated with present knowledge.

For many years the PL was believed to arise from defects, however, evidence has been accumulating that the PL comes from self-trapped
excitons (STEs). As discussed in the next chapter, the defect and STE models are similar in most properties; they differ primarily in that in the defect model the number of excitations is limited to the number of defects, while STEs can exist at any site in the solid. Thus the simple configurational-coordinate diagram in Fig. 10.4 can still be used, simply by interpreting the excited luminescence center as and STE instead of an exciton trapped at a defect. The strongest evidence for STEs is thus the observation (Higashi and Kastner, 1983) that in crystalline $\text{As}_2\text{S}_3$ the PL intensity did not saturate even at excitation densities exceeding $10^{19}\text{cm}^{-3}$.

The initial excitation near the band edge results in a free exciton (FE). This excitation is extended, and therefore does not interact strongly with the lattice. The dependence of the total energy on configuration coordinate $Q$ is thus the same as that in the ground state; for simplicity we assume a harmonic form for the potential. The special features of the model arise from the existence of a second state for the exciton, one that does interact strongly with the lattice, and has a lower minimum energy than the free exciton.

Before the lattice distorts (i.e. for $Q=0$) the STE lies higher in energy than the FE. This need not be the case, but is inferred in the case of c-$\text{As}_2\text{Se}_3$, c-$\text{As}_2\text{S}_3$, and c-Se from the existence of sharp structure in the absorption coefficient near the band edge: absorption directly into the STE would give a broad spectrum, similar in width to the luminescence band.

The figure shows that the higher energy of the STE in the undistorted lattice also implies a barrier to the self-trapping process, which must be overcome either thermally or by tunneling. This barrier
has in fact been observed in c-Se, in which it is manifested as a delay before self-trapping occurs (Chen, Robins, and Kastner, to be published). Those results are in surprisingly good quantitative agreement with purely parabolic configurational-coordinate diagrams.

In the amorphous material, there is no sharp structure in the absorption coefficient, so there may be direct excitation of the STE. Evidence for such direct excitation is provided by the experiments of Murayama, Ninomiya, Suzuki, and Morigaki (1977), and those of Higashi (1982) and Higashi and Kastner (1981, 1983). In studies of the prompt (-10ns) PL in a-As$_2$S$_3$, these authors observed that the PL spectrum shifts in parallel with the excitation energy, for $h\omega_x \lesssim 2.3eV$ (at which $\alpha \sim 30\text{cm}^{-1}$). For excitation above 2.3eV, this constant-Stokes-shift component is joined by low-energy light again, as manifested by the increased linewidth.

This observation, together with a decreased quantum efficiency and a loss of polarization memory for high-energy excitation, led Murayama et al. and Higashi and Kastner to suggest that 2.3eV is just sufficient to create excitons at the exciton mobility edge. Higher-energy excitation results in mobile excitons that can find the lower-energy luminescence centers as well as nonradiative centers. Low-energy excitation directly excites PL centers, as evidenced by the constant Stokes shift.

The PL is similar in crystalline and amorphous chalcogenides; the models we have described seem quite different. The models can be reconciled using a configurational coordinate diagram shown in Fig. 10.5. The different curves for the STEs correspond to a variety of different local environments (inhomogeneous broadening). Low energy excitation excites the low-energy STEs directly—hence the broad absorption band in amor-
Figure 10.5. Possible configurational-coordinate diagram for the disordered material (either amorphous or high-temperature crystal). There are many different energies for STEs due to differences in local environment.

Amorphous materials (and also crystals at room temperature). Higher energy excitation creates free excitons, which are free to diffuse but eventually are self-trapped if they do not first recombine non-radiatively. We will return to this model when we have discussed the photocurrent quantum efficiency.

This figure is identical to that inferred by Murayama, Suzuki, and Ninomiya (1980). Higashi and Kastner implicitly assumed that a distribution of free exciton energies was responsible for the broadening. Presumably both effects are important, but there may be important differences between the two models. A priori, one might expect the
effects of disorder to be greater for localized states; the model shown would then be preferred.

10.5 Transient Photocurrent Quantum Efficiency

Another important indication of the nature of excitations at different energies is the photocurrent carried by the excitations. Traditionally this has been studied by measuring the steady-state photocurrent produced by photons of various energies. For sufficiently low light levels, this current will be directly proportional to the sum of $\eta \mu \tau$ product for the electrons and for holes. Here $\eta = \eta_{PC}$ is the quantum efficiency for creating holes and $\mu$ is the average mobility of the carriers during the time $\tau$ before monomolecular recombination.

Unfortunately, the available signal-to-noise ratio often does not allow measurement in this small-signal limit; one must use higher intensities to get a measurable signal. The recombination time is then shortened by the presence of a high density of photoexcited carriers. In the ideal case, the lifetime is inversely proportional to the excited density, which is proportional to the current, so the current $i$ in this regime is proportional to the square root of intensity $I$ ($\gamma = \frac{1}{2}$ in $i \propto I^{\gamma}$), and if one can measure the dark conductivity the $\eta \mu \tau$ product can be determined.

In amorphous semiconductors, however, the situation is not so clear. For one, $\gamma$ is rarely exactly one half, and the correction for recombination is thus problematic. The problem is particularly severe for high excitation energies, where the short penetration depth requires high excitation densities for measurable signals, and thus large corrections. One must also know accurately the distance over which carriers are distributed; because of diffusion this may be longer than the
penetration depth \( a_x^{-1} \) of the light.

These problems are ameliorated by measuring the transient photocurrent. If the recombination time \( \tau \) is longer than the observation time, then recombination will not affect the resulting current, and no correction is necessary. Furthermore, it can be directly determined whether this condition is satisfied, since recombination will cause a more rapid decay of the current with time.

The TPC measurement yields the product \( \eta \mu(t) \), where \( \mu(t) \) is the mobility at the measurement time \( t \). If the mobility has the same power-law time dependence for all excitation energies (which can be checked directly) then it is reasonable to assume that its magnitude is also the same. The variations in magnitude of the TPC versus \( \bar{\hbar} \omega_x \) then yields the relative variations in \( \eta_{PC} \) (the absolute magnitude remains undetermined).

The results of room temperature TPC measurement on a-As\(_2\)Se\(_3\) are shown in Fig. 10.6. At room temperature, photon energies \( \bar{\hbar} \omega_x \gtrsim 1.9\text{eV} \) require such a small areal density of excitation to avoid recombination (high volume density) that the signal is unmeasurable with the present sensitivity. The data are normalized to the number of absorbed photons; since the sample is 0.5 cm thick, this always differs from the (measured) number of incident photons by less than a factor of three. It should be stressed that it became possible to attain these data only with the continuously tunable pulsed output of the YAG-laser-pumped dye laser/ Raman shifter system.

It is clear that there are two different regimes of behavior: for \( \bar{\hbar} \omega_x \gtrsim 1.5\text{eV} \), \( \eta_{PC} \) is essentially constant, while at lower photon energies it depends more-or-less exponentially on \( \bar{\hbar} \omega_x \). It is intriguing that the
Figure 10.6. Relative room-temperature quantum efficiency $\eta_{PC}$ for a-As$_2$Se$_3$. The plot shows the magnitude of the TPC at 1ms time delay, normalized to the number of absorbed photons. Unity corresponds to $2 \times 10^{-7} (\text{cm}^2/\text{V-s}) \times (\text{e/photon})$.

division between these two regions occurs at essentially the same energy as the division between the exponential edge and low-energy tail in the absorption coefficient, to wit, $\hbar \omega_x = 1.5\text{eV}$. This suggests that the two regions correspond to distinct absorption processes and subsequent ionization processes; that is, that the initial excitations are quite different in the two regions. We will discuss the two regions in turn. Interestingly, measurements of $\eta_{\text{PC}}$ for holes in a-$\text{Si: H}$ (Jackson, Nemanich, and Amer, 1983) gave a similar spectrum.
10.6 Quantum Efficiency in Exponential Absorption Edge

We consider first the "Urbach" process giving the exponential absorption edge. It is striking that the quantum efficiency of this process is constant to within a factor of two while the associated absorption coefficient is changing by more than three orders of magnitude. The constancy of $\eta_{PC}$ contrasts sharply with the strong variation with energy of the PL quantum efficiency: the latter suggests that the excitations produced by different photon energies are very different, but the former shows that any difference does not affect the ionization probability.

This observation would not be surprising if $\eta_{PC}$ were close to 100% at the high temperatures at which PC measurements are possible. If all excitons become ionized, then differences between them may not matter. Indeed, Mort, Chen, Morgan, and Grammatica (1981) estimated the low-field quantum efficiency as 45%, rather close to 100%.

To provide more insight into this question, we can examine the magnitude of the high-photon-energy quantum efficiency at different temperatures. Figure 10.7 shows data for $T=226K$; the correction for fraction of photons absorbed used the measured room temperature absorption coefficient, but the difference should not be large, and is totally irrelevant for high photon energies. Measured as photocurrent at 1ms per absorbed photon, the data at the two temperatures differ by a factor of ~500, but this discrepancy includes not just any change in $\eta_{PC}$, but also the intrinsic temperature dependence of the multiple-trapping process. To correct for this, the low temperature data should be multiplied by a factor of $(\nu_0 T_0)^{\Delta T/T_0} \approx 15$. This still leaves a factor of 30 in magnitude between the two temperatures.

Thus even if $\eta_{PC}$ were close to 100% at room temperature, it appears
Figure 10.7. $n_{pc}$ for $T=226K$. The room temperature absorption coefficient was used to determine the number of absorbed photons. Unity corresponds to $4 \times 10^{-9} \text{cm}^2/\text{V-s} \times \text{e/photon}$.

to be substantially lower at 226K, yet it is still constant throughout the exponential edge. It is possible that the correction for the temperature dependence is in error, for example if the estimates of the temperature dependence of the hopping process given in Chapter 7 are grossly wrong; it is unlikely that this will give a factor of thirty. It is also possible that space-charge effects (see Appendix B) are reducing the current at low temperatures. Any such reduction should depend on the average illumination, however, and the average illumination varies substantially for different photon energies because of the idiosyncrasies
of the laser. The scatter of the data is substantially less than a factor of thirty. Nonetheless, it should be noted that Mort et al. reported "weak" temperature dependence of the magnitude of the quantum efficiency, which suggests that there may indeed be a problem with the magnitudes.

The insensitivity of $\eta_{PC}$ to photon energy is in marked contrast to the well-known case of Selenium, in which the quantum efficiency depends strongly on photon energy. Similar results have been reported for steady-state measurements on a-As$_2$Se$_3$ (Main and Owen, 1973; Robertazzi, 1982), but the reliability of the present measurements suggests that those results, if not those for Se, are in error. The dependence in Se has been modeled using the Onsager model, in which the probability of ionization of an exciton depends on the separation of the electron and hole immediately following their thermalization to the band edge. Extra photon energy is presumed to increase the time that carriers have to thermalize, and thus their eventual radius.

Since in a-As$_2$Se$_3$ (and also in a-As$_2$S$_3$) we see no dependence on the photon energy in the exponential absorption region, it must be that the excess photon energy is incorporated in a way that does not affect the final ionization probability. This could happen if all the energy were quickly dissipated into phonons and all excitations decayed into the same state after a few picoseconds— but this would conflict with the observation from luminescence that the final states are different, unless $\eta_{PC}$ is essentially 100%. This situation would probably change at lower temperatures, since the PL is known to be efficient.

Alternatively, if the excitons consist of an electron in a deep state and a hole weakly bound to the electron, the ionization probabil-
Figure 10.8. Possible model for T-dependent, $\hbar\omega_x$-independent $\eta_{PC}$. If the hole is only weakly bound to a highly localized electron, the rate of ionization into the continuum (shaded) will be independent of the particular state of the electron (Fig. 10.7).

... It will depend only on the escape of the hole. Excess photon energy can be taken up by the electron, and the ionization probability can be independent of $\hbar\omega_x$. Such a picture is in good agreement with the description of the edge as an inhomogeneously-broadened band of self-trapping states, as in Fig. 10.5.
10.8 Quantum Efficiency in Low-Energy Tail

![Diagram showing energy levels]

Figure 10.9. Comparison of energies, showing how surprising it is to get reasonable quantum efficiencies with 0.9eV excitation.

We turn now to the quantum efficiency below about 1.4eV, that is, in the low-energy absorption tail. Since this tail is believed to arise from states deep in the gap, it is not surprising that the $\eta_{PC}$ should be smaller. What is actually surprising is how slowly it drops off. For example, even at $\tilde{\eta}_{0x}=0.9eV$, which is the conductivity activation energy $E_0$, $\eta_{PC}$ is still 0.1% of its high-photon-energy value; there is no evidence of a 26meV ($kT$) wide cutoff corresponding to excitation from the Fermi energy. Fig. 10.9 illustrates the energies involved in this comparison. The results of Thio, Monroe and Kastner (1984) imply that there
are very few states closer than 0.35eV to the Fermi level; any excitation of deep states must come from these. The large quantum efficiency shows that these carriers are released by the observation time \( t = 1 \mu s \), so they must be closer than \( kT \ln(v_o t) \sim 0.5eV \) to the reference energy \( E_o \). \( E_o \) is roughly \( E^* - E_F \), which differs from \( E_o - E_F \) by around 0.1eV. We would therefore have a contradiction if the high \( \eta_{PC} \) continued to 0.85eV; this is almost the case. Electron-phonon coupling would broaden the edge, but it would also move the absorption to higher photon energies. In sum, the large quantum efficiency is very surprising.

One simple approach to modeling \( \eta_{PC} \) is to assume that there are two absorption processes corresponding to the "Urbach" and "Tauc" regions of the spectrum. The Urbach process dominates absorption in the exponential tail, \( \alpha(\bar{\omega}) = \alpha_U(\bar{\omega}) \), and the corresponding quantum efficiency \( \eta_U \) appears to be independent of \( \bar{\omega} \). The Tauc process is the dominant absorption mechanism in the low-energy tail, where \( \alpha(\bar{\omega}) = \alpha_T(\bar{\omega}) \). Clearly the quantum efficiency for this process, \( \eta_T \), is lower. The overall quantum efficiency is just the average for the two processes, weighted by the fraction of the photons each absorbs:

\[
\eta(\bar{\omega}) = \frac{\alpha_U(\bar{\omega}) \eta_U + \alpha_T(\bar{\omega}) \eta_T(\bar{\omega})}{\alpha_U(\bar{\omega}) + \alpha_T(\bar{\omega})}.
\]

(10.3)

The simplest situation would be if \( \eta_T \) were negligible. In this case, we would have \( \eta \approx \alpha_U/\alpha_T \), which would be an exponential with a width of approximately 50 meV. The observed edge is much wider than this, closer to 100 meV. Of course, it may be that the Urbach process does not have a truly exponential absorption. The inferred absorption spectrum is shown in Fig. 10.10; the curve shows a distinct change in slope near 1.4eV.
Figure 10.10. If the Urbach process has \( n = n_U \) and the Tauc process has \( n_T = 0 \), the \( \alpha_U = \alpha_{\text{free}} = \alpha \times n \), shown here. This function is clearly inconsistent with a single exponential.

It therefore seems that the decrease in \( n_{\text{PC}} \) cannot be modeled as simply the competition between two processes with different quantum efficiency; unless the extrapolation of the exponential edge to lower energies is very wrong, we must assume that there is a significant (exponential) energy dependence to the quantum efficiency for the Tauc process itself. This would not be a surprising result; the observation
of an $\hbar \omega$ independent quantum efficiency at higher energies was more surprising.

![Graphs showing $\alpha_\nu$, $\eta_\nu$, and $\eta_{\text{tot}}$ vs. $\hbar \omega$.]

Figure 10.11. Independent, competing absorption processes. If the quantum efficiency for each process is nontrivial, a smooth transition from one to the other is coincidental.
What is difficult to understand is why the transition between the two regimes of quantum should occur have the form it does: the smoothness of the transition requires that $\eta_T = \eta_U$ at 1.4 eV, the same energy where $\alpha_T = \alpha_U$. If $\eta_T$ were substantially lower then one would expect an edge with a 50 meV width to precede the broader exponential. If $\eta_T$ were substantially higher, then the transition should occur at lower energies, perhaps giving a peak. Such a peak has been observed by Robins (1979), by Robertazzi (1982), and by Andrieussens et al. (1983). Certainly the data are not good enough to put particularly strong limits on the near-equality of $\eta_T$ and $\eta_U$ at the crossover (they are within factor of five), but the similarity suggests that the two processes may be related.

![Graph](image)

Figure 10.12. The room temperature TPC quantum efficiency for $a$-$As_2S_3$. 

On the other hand, the near-equality may simply be coincidence. This latter possibility draws some support from similar measurements made on a-As$_2$S$_3$, shown in Fig. 10.12. In this case no change in behavior of $\eta_{\text{pc}}$ is seen as the photon energy passes from the Urbach to the Tauc regions, although the data quality is obviously poor.

### 10.9 Time Dependence at Different Photon Energies.

The interpretation of the magnitude of the photocurrent as reflecting variations in the quantum efficiency depends on the mobility $\mu(t)$ being the same for all photon energies. The current has a nontrivial power-law dependence on time; since this results from a complex multiple-trapping process, it seems reasonable to assume that if the time dependence is the same, then the magnitude must also be the same.

To see if this criterion is satisfied, we have plotted the best-fit power-law exponent for the same data of Fig. 10.6, in Fig. 10.13. This data was not taken with this purpose in mind; as a result the system time constant was slightly too long for accurate time dependences and the slopes are slightly too steep. Nonetheless, the trend is clear: lower photon energies give steeper slopes. Some of the data at the highest photon energies is slightly steeper as a result of recombination; note that this had no substantial influence on the magnitudes in Fig. 10.6. At low photon energies, however, the density should be very low, and recombination therefore negligible. The steeper slopes thus represent a real, if small, effect.

Such a steepening is surprising: the expectation was that the excitation energy could only make the decay slower. The reasoning is illustrated in Fig. 10.14: one might expect that at lower excitation energies the resulting carriers might be left in traps rather than mobile states.
Carriers at a depth $|E|$ will not contribute to the conduction until their release time $v_o^{-1}\exp(|E|/kT)$. That is, they will not participate until the demarcation energy reaches them. Because the number of carriers participating in the conduction is monotonically increasing, the current must decay more slowly than the intrinsic process.

Thus direct excitation cannot be the explanation of the changes, if the high-photon energy data is taken as reflecting the intrinsic time dependence. It seems implausible that the low-energy excitation should represent the intrinsic process, since extensive studies done with high
excitation energies are entirely consistent with the assumption of excitation into the bands. We must therefore look elsewhere for an explanation of this small but intriguing effect.

One possibility is recombination, which is known to make the decay faster. It is difficult to imagine how this might work, since the density of excitations is orders of magnitude smaller for the weakly
penetrating light than for the 1.9eV light, which exhibits bimolecular recombination. The quantum efficiency is low, which makes the density of free carriers even smaller. It is possible that the other (non-photocurrent) entities created by the excitation are enhancing recombination, however. To check this, I tested for the intensity dependence of the slope at low energies (Fig. 10.15). The signal-to-noise is marginal, but the slope is, if anything, steeper for lower intensity light, completely opposite to the effect expected for recombination.

One is therefore forced to consider that possibility that the recombination centers are being saturated at high intensities, so that they speed up the decay only for moderate intensities. The density of
excitations for 10μJ pulse energy is of order $10^{15}$ cm$^{-3}$, although the steady-state density may be substantially higher. The saturation of recombination centers is strongly suggestive of a role for the intrinsic defects discussed in the next chapter.

Another interesting model involves a change in the band-tail width $kT_o$ with excitation density. This would be expected if, for example, the width results in part from the presence of charged defects and the resultant potential fluctuations. If the excitation neutralizes these defects, then the band tail will narrow and a steeper decay will be observed. To explain the intensity dependence, however, one must assume that the defects become charged under illumination (perhaps through photoionization of trapped carriers).

The explanations just discussed are intriguing, but they are quite complicated considering the small size of the effect. Nonetheless, more simple explanations (except, conceivably, experimental artifacts) do not seem to suffice. Given the potential insight into defects or tail widths or both that may arise from further investigation, it may be worthwhile to explore this situation further. Especially promising would be the use of strong bias illumination (see Chapter 12), perhaps provided by a diode laser, in conjunction with low-energy excitation. If this experiment is attempted, some care must be taken with the geometry (see Appendix C).

The observation of a different time dependence seems to invalidate the claim that the magnitude of the current gives an indication of the quantum efficiency. Certainly the change in slope is negligible for excitation above 1.4eV, so the arguments of Section 10.6 still apply. For low-energy excitation, the situation depends on the time dependence
at earlier times than those accessible. If the change in exponent of 
-0.1 continues unchanged back to $10^{-12}$s, the true quantum efficiency 
will be larger by only a factor of ten than that shown in Fig. 10.6. 
Thus the time dependence does not suffice to explain the variation, and 
only makes the extremely weak dependence of $n_{PC}$ on photon energy more 
dramatic.
CHAPTER 11

INTRINSIC DEFECTS AND SELF-TRAPPED EXCITATIONS

In this chapter we finally introduce the subject of states deep in the gap, as opposed to band-tail states. Such states are believed to be responsible for the limited usefulness of chalcogenide semiconductors, so if one were able to modify their density it would have tremendous technological impact. The intrinsic defect models discussed here suggest that such modification may not be possible. This model has been based largely on negative results, and despite a decade of work the evidence is still sparse. It is therefore important to learn as much as possible from the experiments that we have. Some of the observations traditionally ascribed to defects may reflect the self-trapping of excitations; these topics are closely related.

11.1 Negative U

Traditionally, the existence of a large density of states near the Fermi level has several experimental signatures. These include paramagnetism (including spin resonance) from unpaired spins, large low-temperature dc conductivity from carriers hopping near the Fermi level, poor transparency for sub-band-gap light, and pinning of the Fermi level as deduced from doping and field-effect measurements. In chalcogenide glasses, all of these effects are strikingly absent, except for the Fermi-level pinning. For example, esr measurements yield an upper limit of $10^{15}-10^{16}\text{cm}^{-3}$ unpaired spins. In contrast, in order to explain the small size of the field effect as a bulk phenomenon one must invoke densities of $10^{17}-10^{18}\text{cm}^{-3}$.
This apparent contradiction was eventually resolved by Anderson (1975), extending his concept of a gap in the one-electron DOS induced by strong electron-phonon coupling (Anderson (1972)). If such coupling is strong enough there can exist an effective attraction between electrons, since the distortion of the lattice by the first electron may lower the energy of a second electron enough to overcome their Coulomb repulsion. This attraction is described by a "negative Hubbard U," which is the energy difference between having two electrons on the same rather than different sites. As a result, although there is a density of two-electron states sufficient to pin the Fermi level, there is a gap in the one-electron DOS, since extra energy U must be supplied to unpair electrons. Anderson envisioned that the negative U acted at all sites in the solid, and went on (1976) to propose that the observed gap of amorphous semiconductors simply reflects the negative U, a view endorsed by Lio- ciardello Stein, and Haldane (1982). The mainstream of the amorphous community, however, has taken a different course.

11.2 Intrinsic Defects

This course was initiated by Street and Mott (1975) and Mott, Davis, and Street (1975), who proposed that the negative U acted only at intrinsic structural defects, which are present in relatively small density. They specifically mentioned dangling bonds in the network as possible defects, but this does not explain why only the Chalcogenides show the apparent discrepancy requiring negative U. Kastner, Adler and Fritzscbe (1975) pointed out the importance of the chalcogen lone-pair electrons, which are thought to be easily available for chemical reactions. These reactions tend to result in doubly-occupied bonds (no unpaired spins), and can stabilize charged defects as compared with
uncharged defects having unpaired spins. Interestingly, the low coordination number of the chalcogen atoms, which contributes to their glass forming tendency, also allows for greater relaxation in response to a change in electronic distribution.

One expects defects to be present in a liquid with a density determined by their (free) energy of formation $E_{\text{form}}$: $N(T) = N_0 \exp(-E_{\text{form}}/kT)$. Here $N_0$ is the total number of sites at which such defects might form, and is of the same order as the number of atoms. On passing below the glass transition temperature $T_g$, structural relaxations become very slow and defects in excess of the equilibrium density can no longer be removed. The density of defects is then simply the number that were present in the liquid near $T_g$.

This prediction of the intrinsic defect models was supported by the annealing studies of Abkowitz and Pai (1977) and Abkowitz, Pochan, and Pochan (1980) on a-Se. It has recently received numerical validation by the monomolecular-recombination measurements of Thio, Monroe and Kastner (1984) mentioned in Chapter 4. Fig. 11.1 shows the density of thermally occupied recombination centers inferred from those measurements. (The possible implications of the hopping process for this interpretation have been only partially explored.) The higher activation energy at high temperatures is associated with thermal generation of intrinsic defects; this assignment is supported by the fact that the extrapolated infinite-temperature density is of order $10^{21}$ cm$^{-3}$.

What relaxations become slow near $T_g$? Certainly large-scale rearrangements associated that allow the material to flow in response to shear stress become slow, as reflected in the viscosity. In addition, the shorter-scale relaxations associated with crystallization become
Figure 11.1. Inferred Density of Recombination Centers in a-As$_2$Se$_3$. The larger activation energy above $T_g$ is taken as evidence for thermal generation of defects.

slow; these presumably involve breaking and reforming of large numbers of bonds. Any diffusion of atoms will likewise be retarded. Purely local reactions, however, may require a somewhat higher energy because of the added stiffness of the matrix, but should not be "frozen out" in the same sense. This suggests that defects that are truly frozen-in must require a major rearrangement to disappear. For example, a dangling bond, with all nearby atoms properly bonded, cannot be removed without disrupting the neighbors coordination. Any defect that has conserved
quantum number (other than charge) different from the bulk will have
difficulty relaxing below $T_g$.

11.3 Elemental Chalcogenides (Selenium)

The prototypical material for which the negative-$U$ arguments were
developed is amorphous Selenium, because of its simplicity. The low-
lying defects were believed to have the form of charged singly and tri-
ply coordinated selenium atoms (Valence Alternation Pairs or VAP's; see
Kastner, Adler and Fritzscbe 1975) as shown in Fig. 11.2. Detailed cal-
culations by Vanderbilt and Joannopoulos (1979, 1983a,b,c), however,
indicated that this material should not have a negative $U$, and that the
lowest-lying (most abundant) defects should have free spins. The reason
for this is that the neutral defects in Selenium also have a low
energy because of interaction between the p-orbital on the dangling Se
atom and the $\pi$-orbital on the Se atom to which it is bonded. This
mechanism is not applicable to $\text{As}_2\text{Se}_3$, since the chalcogens are bonded
to Arsenic atoms, so this result is not necessarily in conflict with the
apparent negative $U$ in $\text{As}_2\text{Se}_3$. Indeed, the calculations supported the
notion that structural relaxations are important and have energies com-
parable to the electronic energies.

These arguments thus make it plausible that pairs of oppositely
charged defects should have a lower energy than pairs of neutral
defects. They do not, however, address the important question of why the
density of neutral defects is so small. As noted before, except for the
apparent Fermi-level pinning, all the normal indications of large gap-
state densities give negative results in the chalcogenides. Except for
the negative-$U$ defects, then, the gaps of these materials are extraordi-
narily clean. If one could somehow passivate the negative-$U$ defects,
the chalcogenide glasses could be very useful for applications requiring doping, as they already are for photoconductive applications.

11.4 Compound Glasses

Further complications arise in the defect chemistry of compound glasses such as As$_2$Se$_3$, as discussed by Kastner (1978) and by Vanderbilt and Joanneopoulos (1981). The availability of different species of atoms allows the possibility of a great variety of defects, since there may now be like-atom bonds (LABs) as well as mal-coordinated atoms (MCAs).
Vanderbilt and Joannopoulos speculate that the LAB's may have a low formation energy in glassy As$_2$Se$_3$, if there is no strong intermediate-range order. They will therefore be present in large densities ($\sim 10^{19}$ cm$^{-3}$) and can easily accommodate any stoichiometry variations. They may give rise not to deep states but only to shallow, hydrogenic traps. In the crystal, by contrast, LAB's are unlikely: simple substitutions cause much less disruption of the lattice. The difference in density of LAB's may help explain why only holes are mobile in a-As$_2$Se$_3$ (Pfister and Scher, 1977) while only electrons are mobile in c-As$_2$Se$_3$ (Marshall, 1977).

Other defects may give states in the gap similar to those attributed to VAPs in a-Se. If they are to pin the Fermi level, it must be possible for two types of defect with different charge (both spinless) to interconvert. The potential defect reactions are thus constrained by the need to preserve the number of atoms of each species. A defect such as a Se-Se LAB, for example, necessarily involves two fewer Se-As bonds than a Se-As-Se bridge. Preserving the stoichiometry requires other defects as well. To help with bookkeeping on such reactions, Vanderbilt and Joannopoulos introduced the quantum number "Chalcogen excess". For example, the Se-Se LAB has a chalcogen excess of one, since it can be created by adding one chalcogen to stoichiometric material. Any defect reactions must conserve the total chalcogen excess of the material.

An additional constraint is imposed if the defect reactions are supposed to pin the Fermi level even in the solid state. Such a reaction can involve the transfer of electrons, but presumably not of atoms, over long distances. The lack of atomic mobility in the solid requires than any such reactions proceed through a local rearrangement of exist-
**Figure 11.3.** Simplest probable Fermi-level pinning reactions in a-As$_2$Se$_3$. The important role of the lone-pair p-orbitals in the reactions is emphasized.
preserved in the reaction. Even for elemental glasses it may be important, however. For example, a singly coordinated Se in the midst of continuous chains cannot disappear: there will always be either a singly- or triply-coordinated defect. There may be a quantum number ("bond excess"?) which quantifies this constraint as well.

11.5 One-electron Density of States

Although the specific defects responsible are not well determined, the idea that intrinsic (as opposed to impurity-related) defects are the most important states in the gap has come to be widely accepted. The simplest models with a negative U contain defects that are positively and negatively charged in the ground state; to avoid bias in favor of any particular defects let us denote them \( D^+ \) and \( D^- \). The Coulomb interaction between these defects (which is poorly screened in these highly insulating materials) results in a highly unusual spatial distribution of defects.

In the absence of the Coulomb interaction, there would be a "free" density \( N_f = N_0 \exp\left(-E_{\text{form}}/kT_g\right) \) of defects. These defects would be distributed at random and thus be described by the nearest-neighbor distribution function \( 4\pi N_f R^2 \exp\left(-\frac{4\pi}{3}N_f R^3\right) \) shown in Fig. 11.4(a). Since these defects must be created in pairs, the formation energy \( E_{\text{form}} \) is just half of the energy to form a pair.

When the Coulomb interaction is included, there will be an enhanced probability of finding close pairs of opposite charge (and a diminished probability for pairs of like charge), because the extra Coulomb energy will decrease (increase) the formation energy of the pair. Here we focus on the oppositely-charged pairs. If the spatial distribution of the pairs comes into equilibrium at a temperature \( T_g \), the probability of
Figure 11.4 (a) Nearest-neighbor distribution function for uncorrelated pairs. (b) Enhancement of density by Coulomb interaction. (c) Resulting distribution function has two peaks, one at the average distance $-N_f^{-1/3}$, and one at the minimum distance $R_{\text{min}}$. 
finding a pair at separation $R$ will be enhanced by a Boltzmann factor \( \exp(-e^2/\varepsilon R k T) = \exp(-R_C/R) \), where $R_C$ is the Onsager radius at $T_g$ (see Fig. 11.4(b)). The enhancement of the density will will be important for pairs closer than $R_C$. For $T_g=550K$ and $\varepsilon=10$, the Onsager radius is about 30Å.

The modified distribution is shown in Fig. 11.4(c). The number of distant centers is just the free density $N_f$. The number of correlated centers ("intimate" pairs, in a loose sense), depends on the minimum distance $R_{\text{min}}$ to which pairs can approach; if $R_{\text{min}}$ were zero there would be infinitely many such pairs. In practice, $R_{\text{min}}$ is the closest that defects can get without being able to annihilate one another, and is probably of order 5Å. The number of close centers differs from $N_f$ by a factor of order $N_f R_{\text{min}}^3 \exp(-R_C/R_{\text{min}})$.

The one-electron DOS implied by this discussion is sketched in Fig. 11.5: for reasonably low $N_f$, there is still a peak corresponding to isolated defects; it is the energy from the Fermi level to these peaks which is presumably measured by the lower activation energy of Thio et al. In addition there is a broad, roughly exponential manifold of states extending towards the bands (Orenstein, 1981; Nagy, 1984). The extent of the defect band is $e^2/\varepsilon R_{\text{min}}$, which may be several tenths of an eV.

In this model the change in the one-electron energy resulting from the Coulomb interaction is the same as the change in the formation energy. This explains the exponential DOS in a natural way, especially since the width of the exponential $kT_o$ for holes in a-As$_2$Se$_3$ is similar to the thermal energy $kT_g$ at the glass-transition temperature. In spite of its attractiveness, however, there are important difficulties with this interpretation. The DOS is exponential over at least 0.5eV (see
Figure 11.5. Expected one-electron DOS resulting from Coulomb correlations. Note the structure corresponding to preferred near-neighbor distances.

Chapter 4); one would expect to see deviations from the perfect exponential, because of the $R^2$ factor and also because at small separations (comparable to interatomic spacings), some distances are more likely than others. In $a$-$As_2S_3$ the exponential is about twice as wide as in $a$-$As_2Se_3$, although $T_g$ is similar. In addition, the exponential DOS has been observed to be different for electrons and holes in $a$-$Si:H$; the very existence of an exponential DOS in $a$-$Si:H$ is problematic, since this material is not in equilibrium at any temperature.

Nonetheless, the distribution of energy levels resulting from electrostatic potential fluctuations must be present if the intrinsic
defects are charged and are present in the densities estimated. Any model for the DOS involving charged defects should include these effects. The energy distribution may be important to the dynamics of the apparently defect-related metastable effects, which we now discuss.

11.6 Low-temperature Long-lived Effects

Intrinsic negative-U defects were originally introduced to reconcile the apparent Fermi-level pinning with the absence of the normal signatures of a high DOS at the Fermi level. Since their introduction, many other observations in chalcogenide glasses have been attributed to the excitation of intrinsic defects. In particular, several correlated effects are observed as long-lived changes in properties following illumination with Urbach-tail light at low temperatures.

For example, although the system is diamagnetic in its ground state, illumination produces a spin signal, presumably reflecting a conversion of the spin-paired ground state of the defects to a long-lived excited state. Bishop, Strom, and Taylor (1975, 1976, 1977) (BST) identified a Chalcogenide-associated center and an "Arsenic-associated" center by analysis of the EPR lineshape. While the uniqueness of their identification of the Chalcogen center as a hole and the As center as an electron is questionable, it is unambiguous that there are highly localized electronic states induced by the low-temperature illumination.

Another effect associated with low-temperature illumination is the fatigue of the luminescence intensity. Traditionally this effect has been ascribed to the conversion of some fraction of the luminescence centers into non-radiative centers during the radiative recombination process. In addition to this straightforward explanation, however, there are other possibilities, such as Auger process in which the
recombination energy is transferred to a nearby trapped carrier. Higashi and Kastner (private communication) have also observed, in a-As$_2$S$_3$, an increase in the absorption of the excitation light with time; the competition of the new absorption process with absorption by luminescence centers may play a significant role in the fatigue as well. This phenomenon is presently under study by T. Thio.

Finally, and most relevant to our work here, Cernogora, Mollot, and Benoit a la Guillaume (1973) and later BST observed a broad, long-lived absorption band extending from roughly 0.7eV up to the band edge of the sample. The spectrum of this "metastable" PA band is shown in Fig. 11.6 on the same scale as used for later transient PA figures: the spectrum is extraordinarily flat above threshold at $\hbar\omega > 0.7$eV.

These effects are correlated in the sense that they seem to grow together on illumination at low temperature, and they are all generated most efficiently by light with an absorption coefficient of around 100cm$^{-1}$ (the same light which generates PL most efficiently). In addition, all these long-lived excitations can be removed by thermal annealing, or by "bleaching" with low energy light (-1eV). It is difficult to know how much confidence to place in these correlations, however, since spurious correlations are possible— for example, a separated electron and hole will give separate but correlated signals.

Unfortunately, because these long-lived excitations are very time consuming to measure, a precise description of the spectral properties of the generation and bleaching of these centers has been lacking. By using transient measurements we can obtain this sort of information much faster, and some of these experiments can now be performed. First, however, we must provide a means for controlling the long-lived excita-
Figure 11.6. "Metastable" Induced Absorption Band in a-As$_2$Se$_3$ (Bishop, Strom and Taylor, 1975).

11.7 Self-trapped States

A longstanding controversy in amorphous semiconductors is the role of polarons, or self-trapped carriers, in the observed properties. In this section we will outline some of the similarities and differences between the polaron picture and the defect picture.

In both the defect and polaron models, the final state is a strongly electron-phonon coupled state deep in the gap. The relevant question is: which phenomenon can be considered primary? Do electrons localized by disorder experience a strong electron-phonon coupling, or
do electrons localized by electron-phonon coupling experience a variety of local potentials? Clearly the distinction is partly semantic.

The polaron bandwidth in a perfect crystal is very narrow, and it would seem that a small amount of disorder should localize the states in this band. Since the electronic description starts with localized states, it seems more fundamental. On the other hand, the electron-phonon coupling results in shifts of the energies of order eV, which is a large fraction of the gap. Since the polaron description starts with electron-phonon coupling, it seems more fundamental.

The primary objective distinction between the defect and polaron picture is the available density of sites. The density of defects must be fairly small, or they would not be considered defects; estimates based on Fermi level pinning are consistent with the observations of Thio et al. of a density of a few times $10^{17}$ cm$^{-3}$ defects. A similar estimate was obtained by Bishop, Strom, and Taylor for the "saturated" density of metastable centers; it may be, however, that this is not an accurate measure of the density of available centers—see Chapters 12 and 13. In the polaron picture (or in the Anderson/Licciardello picture) all sites in the solid are presumably available, so the density should be of order $10^{22}$ cm$^{-3}$.

Other than the density, there seems to be little reason to prefer one or the other model for the deep states. Indeed, there are great similarities between the reactions involved in formation of intrinsic defects and those involved in self-trapping, as we will illustrate with an example. For concreteness, I have chosen a section of the crystal-line form of As$_2$Se$_3$, illustrated in Fig 11.7. The lone-pair orbitals on the Se atoms are shown, since these are believed to play an important
Figure 11.7. A section of the $\text{As}_2\text{Se}_3$ crystal shown in Fig. 1.2, with the lone-pair orbitals on two of the Se atoms shown. The two pieces happen to be part of two separate layers.

role in the stabilization of defects in Chalcogenide semiconductors.

In Fig. 11.8, we have introduced an electronic excitation resulting in structural relaxation. The structure now contains a self-trapped exciton. Our main point is that the resultant species are, at this level of description, the same chemical entity discussed in the defect models. To emphasize this, the chemical representation on the right in Fig. 11.8 identifies the species according to the notation of Kastner, Adler, and Fritzsch (1976): the "C" stands for chalcogen, the subscript represents the coordination number, and the superscript the charge state.
Figure 11.8. The same set of atoms as in Fig. 11.7 after a chemical rearrangement involving an electron-hole excitation. The arrow indicates an Arsenic atom which undergoes a large rearrangement during such a reaction, participating cooperatively in both the electron and hole trapping.

Clearly the species resulting from self-trapping have a great deal in common with the possible Fermi-level pinning defects. In addition all conserved quantum numbers must sum to zero. To the extent that the energy to create an isolated pair is low, the energy to create the STE is lower by the Coulomb energy. There is also an additional 2eV or so available in the self-trapping case, because the electron and hole are on opposite sides of the gap.

The example shown was chosen to illustrate another point. The exciton may self-trap even though neither the electron nor the hole
individually do so; the distortions involved in the reaction shown are cooperative. In more typical cases where STEs are observed, for example, Alkali Halides, one carrier self-traps and the other is weakly bound, as suggested in the last chapter to explain the TPC quantum efficiency.
CHAPTER 12

BACKGROUND EFFECTS AND BIAS BEAMS

The low temperature TPA measurement offers several exciting possibilities, including more detailed description of the metastable induced absorption, examination of the excited state of the luminescence centers, and a better understanding of the hopping-down thermalization. At low temperatures, however, the long lifetime of the excitations causes important complications, since the carriers excited by a pulse can interact with the background. These effects complicate the interpretation of simple TPA measurements. Like many complications, however, this interaction can be turned to our advantage, and by deliberately manipulating the background with bias beams, we obtain an additional probe of the excitations. Such a procedure requires a clear understanding of the effects of the background; the development of such an understanding is the purpose of this chapter. In the process we will explore two traditional classifications of the phenomena: into transient and metastable, and into defect- and tail-related. We will show that neither of these divisions is as clear as it seems.

12.1 Introduction to Background Effects

The normal concept of transient experiments is that the excitation pulse is incident on a sample in its thermodynamic ground state at the temperature of the measurement. In practice, however, the sample is exposed to a continuous series of pulses; in addition, for a TPA measurement the sample is also exposed to a probe beam. As a result, the experiments is not really performed on the ground state, but on the
excited state of the system resulting from the average illumination. Furthermore, if the background is changing slowly with time, it may interfere with the reproducibility of the transient measurements. Actually, nonreproducibility is good, since it serves as a warning of uncontrolled variables that are important enough to affect the results. Indeed, the present exploration of these effects was stimulated by irreproducibility in TPA spectra.

It should be emphasized that, although several of the experiments described here involve the deliberate manipulation of the background using a bias beam, the potential buildup of a background is inherent in repetitively-pulsed transient measurements, as well as in optically probed measurements. An appreciation of the manifestations of such effects is therefore important to a large variety of measurements.

As a crude estimate of the importance of background effects in repetitively-pulsed measurements, let us consider the simplest situation possible, in which carriers are created by the pulsed excitation, survive for a lifetime $\tau$, and then recombine without interacting with other carriers. The signal at any given time is just the sum of the response to all preceding pulses; as a result, the background to which each pulse is added will grow with time until the lifetime $\tau$, at which point carriers will be recombining, on average, as fast as they are being created. This is illustrated in Fig. 12.1.

Once this steady state is reached, as many carriers recombine from the background as are created by the new pulse, so that, just before a pulse, the number of carriers has returned to the value it had just before the preceding pulse. Thus the "truncation error" associated with taking the pretrigger signal as the baseline is not an artifact, but a
Figure 12.1 Build-up of background in noninteracting case.
the background grows until its decay between pulses just
compensates the extra signal from the pulse.

A legitimate reflection of the decay of the background occupation between
pulses. This does not change the fact that it is difficult to interpret
data taken close to the repetition time. This difficulty applies for all
times at which the signal is comparable to the decay of the background,
and is thus more important at shorter times for slow decays.

It is instructive to estimate the magnitude of this background as
compared with the single-pulse response. If \( N_p \) carriers are introduced
per pulse, then in steady state \( N_p \) carriers must recombine by the time
of the next pulse, a time \( t_{\text{rep}} \) later. If the lifetime is longer than \( t_{\text{rep}} \) then these recombining carriers come out of the background, which is decaying at a rate \( N_B/\tau \), where \( N_B \) is the number of carriers in the background; thus \( N_B/\tau = \tau/t_{\text{rep}} \). If the background is recombining with more complicated kinetics (e.g. bimolecular), one should use the appropriate lifetime in the presence of the background, according to
\[
dN_B/dt = N_B/\tau.
\]

This simple calculation demonstrates that one can expect to have a significantly more carriers in the background than are introduced by the pulse when the lifetime of excitations is much longer than the time between pulses. These effects will be particularly important, therefore, at low temperatures (when the lifetimes are long) and when the repetition rate is high (as in many picosecond experiments). Whether the large background affects the results of transient measurements depends on the system. Usually it depends more on the average power of the laser than on the repetition time, because a higher density is necessary to cause important effects on a shorter time scale.

12.2 Background in the Presence of Several States

In the example just discussed, there was only one type of center present, and this was directly monitored by the experiment. In practice, there may be several species affected by the average illumination, and so of these may not show up directly in the measurement, although they affect the resulting signal.

The absence of background effects can be rigorously deduced only by performing the experiment several times, with different average powers but the same pulse energy, that is, with different repetition rates. For a TPA measurement one should also vary the probe intensity to be sure
that the probe is not perturbing the sample. Such experimental checks should be performed periodically to insure the accuracy of the data. Unfortunately, the alteration of the intensity of either these beam usually involves a loss of signal-to-noise ratio, since they are generally used with maximum intensity. The introduction of a third beam, the bias beam, allows the experimenter to demonstrate the presence of background effects without sacrificing signal-to-noise.

12.3 Inducing and Bleaching

As mentioned in the previous chapter, illumination of chalcogenide glasses with bandgap light produces sub-bandgap absorption that persists for very long times at low temperatures (Bishop, Strom, and Taylor, 1975, 1976, 1977). This induced absorption has traditionally been associated with the photo-induced occupation of existing defects, although photocreation of defects is also possible (Biegelson and Street, 1980). Correlated with the growth of the midgap absorption are the creation of an electron paramagnetic resonance (EPR) signal and the fatigue (reduced intensity) of the luminescence. We will refer to light with photon energies greater than about 1.6eV as "inducing" light, although this description should be regarded with caution, as discussed below.

At low temperatures the decay of the induced absorption is very slow. It can be speeded up by raising the sample to higher temperatures, or by illuminating the sample with light within the induced absorption band. Light with $\hbar \omega > 1.4eV$ will be referred to as "bleaching" light. Interestingly, the entire induced absorption band is bleached by illumination at one photon energy. This suggests that the width of the band arises from homogeneous broadening, that is, that each induced centers produces the same spectrum, similar to that spectrum.
It is important to note that the "inducing" and "bleaching" tendency of light with a particular photon energy is probably a question of degree, and depends on the density of centers already induced. For example, although exposing a virgin sample to "inducing" light will increase the absorption, the same light shining on a sample with a well-developed induced absorption may well have a bleaching effect.

The possibility of self-bleaching by the inducing light radically alters the interpretation of the saturation of photo-induced effects observed by BST. The "saturated" value may well represent a balance between inducing and bleaching, rather than a true saturation of all defects present. This possibility is made more plausible by the observation that the induced absorption coefficient of \(10\text{cm}^{-1}\) is comparable to the virgin sample's absorption coefficient of \(-100\text{cm}^{-1}\), so the density of re-excitations is approaching the density of primary excitations. This self-bleaching is virtually certain in the measurements of Bishop, Freitas, and Strom (1984): these authors used \(1.18\text{eV}\) light to induce similar paramagnetic centers to those induced by high-energy light. The saturated densities were smaller by several orders of magnitude, as was the absorption coefficient.

The ability of light to either decrease or increase the absorption is illustrated in Fig. 12.2. The \(1.3\text{eV}\) bias light causes bleaching in a well-darkened sample, but when the sample has been well bleached by \(1.1\text{eV}\) light, it gets darker during the \(1.3\text{eV}\) exposure. This particular data is not conclusive, however, since the \(1.4\text{eV}\) light (which darkens the sample) is present the whole time, so that one could argue that the \(1.3\text{eV}\) bias is simply not as effective at bleaching the excitations caused by the \(1.4\text{eV}\) light. Nonetheless, it is clear that the eventual
transparency results from a competition between the various beams, and these are not very different in photon energy.

Further insight into the inducing and bleaching process comes from the experiments of Orenstein and Kastner (Kastner, 1980). They showed, by simultaneous inducing and bleaching, that the bleaching process was bimolecular, that is, proportional to the square of the induced optical density. This suggests that the bleaching process is simply recombination via photo-assisted diffusion; the probability of recombination
occurring before retrapping is proportional to the density of centers.

In the presence of inducing bias (whether intentional or not) there is a significant density of excitations around. We have already discussed, in Section 4.2, the possible electronic and excitonic absorption processes which can contribute to the observed TPA signal. The description of these processes does not change when one includes background effects; what changes is the effective ground state in which these processes can take place. When background effects are included, one can no longer interpret the excited electrons and holes of Fig. 4.1 simply as excitations above or below the Fermi level; they must be interpreted as increases or decreases with respect to the steady-state background occupation of the states. As a result, bias light can manifest itself in a variety of ways, including enhanced recombination, saturation, and re-excitation.

12.4 Metastability

The long-lived properties induced by bandgap illumination at low temperatures (absorption, EPR, fatigue) are commonly referred to as "metastable." In an operational sense, this means that the effects last longer than the experimenter is willing to wait—normally a time between minutes and a few hours. It is important, however, to distinguish between truly metastable effects—those which decay only on long time scales—and apparently metastable effects—which exhibit a broad distribution of decay times that includes very long times. The distinction is illustrated in Fig. 12.3. In the former case, all of the relevant rates are slow, and no information can be gleaned from time dependence measurements. In the latter case, which is exemplified by power-law time decays, there is decay on all time scales. Transient measurements then
provide information about the same distribution of rates (albeit a different part of that distribution) that is responsible for the apparent metastability.

![Diagram of log effect vs. log time]

Figure 12.3. (a) True metastability: the decay occurs only at long times. (b) Apparent metastability: although much of the signal remains for very long times, the decay at shorter times reflects the same basic process.

Unfortunately, this distinction has not been appreciated by many experimenters, so it is often difficult to tell from a description of a phenomenon to which class it belongs. I will therefore use the term "long-lived" to avoid the implications associated with "metastable." With respect to the phenomenon of induced absorption, it appears that it may well be only apparently metastable. Support for this assertion comes from the fact that significant restoration of the transmission is
observed on a time scale of minutes, after which the decay becomes pain-
fully slow. In addition, the observation in pulsed measurements of
power-law decays suggests the likelihood that some decay also occurs at
longer times.

Further evidence for a distribution of rates comes from annealing
studies of the photoluminescence fatigue by Cernogora, Mollot, and
Benoit al Guillaume (1974). After inducing the fatigue at low tempera-
tures, they observed that isochronal (equal-time) anneals at succes-
sively higher temperatures removed only a part of the fatigue, as shown
in Fig. 12.4. This observation strongly suggests that a distribution of
activation energies is being sampled. At fixed temperature, this distri-
bution is probably manifested as a distribution of decay times.

We may therefore be able to think of the transient measurements and
the "metastable" measurements as accessing different regimes of a single
fundamental process. In this context it is less shocking that the
induced absorption spectrum is similar on picosecond (Fork, Shank,
Glass, Migus, Bosch, and Shah, 1980), microsecond (discussed here and by
Orenstein, 1981), and several minute (Bishop, Strom, Taylor) time
scales.

12.5 Unifications?

The possibility that all these measurements are accessing the same
set of states is intriguing, since the long-lived effects have tradi-
tionally been described in terms of defects, while the microsecond data
seem to reflect the photoionization of carriers in band-tail states (see
Chapter 5). This latter interpretation may be strengthened slightly by
the observation of a broad range of activation energies for the
"defect-related" processes. In view of these observations it is
Figure 12.4. Results of isochronal anneals at various temperatures on luminescence intensity (Cernogora et al. 1974). G is the luminescence intensity at 4K, normalized to the unfatigued (or completely annealed) intensity.

pertinent to ask whether there is any evidence that these supposedly different processes are in fact distinct. Although it would be an overstatement to claim they are identical (see Bosch, 1982 and the response by Orenstein and Kastner), there is no clear dividing line between the two processes.

The reader may recall the discussion of the Coulomb broadening of defect energies presented in Section 11.3. If the band tails indeed arose from such broadened defects, many puzzles would be explained. As we have already noted, however, there seems to be significant circumstantial evidence against that possibility.
An alternative possibility is that the supposedly defect-related properties actually reflect band-tail states. This possibility is not in conflict with the observed saturation of the long-lived excitations, since it is quite likely that this saturation comes from self-bleaching by the excitation light. In addition, the traditional "metastability" of the effects is probably an overstatement of the experimental case. This possibility does appear to conflict with the observations of Thio et al., however. Nonetheless, the absence of true defect effects should be considered seriously; it is important to try to create a unified picture of the many interesting experiments on photoexcited states. Further insight into the possibilities will be gained in the next chapter.
CHAPTER 13

SURVEY OF TRANSIENT PHOTO-INDUCED ABSORPTION EXPERIMENTS

The TPA spectrum varies strongly with excitation energy and temperature; in this chapter we present an overview of these variations, with the intent of identifying different processes contributing to the signal. An identification of the contributions can help to clarify the influence of the various parameters on the excitations, as well. While this chapter represents only a partial survey, we have identified several important features which should be interesting subjects for future study.

13.1 Introduction - Multicomponent Spectra

The TPA results described in Chapter 5 seem to fit in well with the multiple-trapping model. Indeed, these measurements can be regarded as the most direct evidence for the thermalization of the carriers associated with MT. The low-temperature results are also consistent with the model of hopping down thermalization discussed in Chapters 7 and 8.

In the absence of a detailed understanding of the spectrum of the TPA, however, it is unclear how strong these interpretations are. The shape of the spectrum may also give more detailed information about the wavefunction. Such an understanding is not at all easy, as discussed in Chapter 5. Part of the reason for this may be that several different processes contribute to the observed spectrum; if so, one can hope that these processes will behave differently with temperature, time delay, and excitation energy. In general this approach will be only partially successful. For example, the number of excited electrons and holes is
always equal, so if the associated spectra occupy the same spectral region it will be very difficult to separate them. Nonetheless, we will present a preliminary identification of four different contributions to the TPA spectrum in $\text{a-As}_2\text{Se}_3$ on the basis of the present considerations.

We will present data on the induced absorption as a function of several parameters: the excitation energy $\tilde{\omega}_x$ determines the initial state or states created; the temperature $T$ is important in determining their subsequent evolution; the observation time $t$ determines how far this evolution is allowed to progress; and the spectrum in probe energy $\tilde{\omega}_p$ provides more detailed information about the nature of the excitation. In addition, the pulse energy $E^*$ and the presence and nature of any bias light may affect the interactions of excited carriers with other carriers that may be present. Clearly it is impossible to make measurements for every combination of parameters; even if it were possible it would be difficult to interpret.

Because of the large number of parameters, any given plot will represent only a slice through parameter space. The most useful slice for separating spectral components is a spectrum versus probe energy, at fixed time and excitation conditions. It is useful to include an indication of the time dependence as well; this will be represented by an arrow in the direction of the evolution, with a length equal to the decay of the data (fit to a power law) in one decade. Such a plot suppresses any additional information in the time decay, which is important for some of the data.

If the response at some probe energy $\tilde{\omega}_p$ is the sum of the response of two separate components,

$$\Delta \sigma(\tilde{\omega}_p) = \Delta \sigma_1(\tilde{\omega}_p) + \Delta \sigma_2(\tilde{\omega}_p),$$  \hspace{1cm} (13.1)
then the power-law exponent \( s \) is related to the exponents \( s_1 \) and \( s_2 \) of the components by

\[
\frac{d\Delta_0}{d\Delta t} = s - s_1^+ + (s_2 - s_1) \frac{\Delta_0}{\Delta_0_1},
\]

assuming \( \Delta_0 < \Delta_0_1 \). If the spectrum for an individual component decays in the same way at all probe energies (a reasonable definition of component), and if different components decay differently, then the time dependence can be used as a clue to help separate the components. Similar expressions of course apply to variations of the spectra with other parameters, but they are generally not so directly measured.

13.2 Ideal TPA Measurement

The TPA data are plotted as a change in cross section \( \Delta \sigma \) per absorbed photon. For this quantity to have a simple interpretation, several criteria must be satisfied.

The most obvious criterion is that the probe be nonperturbing, that is, a true probe: the presence of the probe should not affect the measured cross section. Since the probe is a requisite for measuring the cross section, this means in practice that the result must be independent of the intensity of the probe, a condition that can easily be checked experimentally. If the probe intensity is found to be too high, it can be reduced by monochromatizing it before the sample (now standard practice, described in Chapter 2), by adding neutral density filters, or by gating the probe so that it only hits the sample during the time data is being collected. [A Gating Circuit has been constructed, which supplies signals to drive the Uniblitz shutter and also to gate the Ithaco preamp, for precisely this purpose.] If the probe beam cannot be reduced to a negligible level, it must be considered part of the bias.
illumination, discussed below.

A second requirement for \( \Delta \alpha \) to be meaningful is that the signal be proportional to the excitation pulse energy \( E^* \), so that the signal can be interpreted as arising from independent single excitations. Naturally one can only observe excitations in proportion to the quantum efficiency for their creation, but a signal that is not proportional to the pulse energy implies interaction, probably recombination, between different excitations. Although the resultant signal can still be useful, the intensity dependence must be an important part of the interpretation.

Another possible source of nonlinearity is an interaction between the pulsed light and the associated background, which should be considered part of the biasing illumination. Note that a dependence of the signal on pulse intensity may actually result from interaction with unintentional bias produced by the pulses.

In order that the effect of bias light be well defined, it must also satisfy certain criteria. If the intent of the bias is to bleach excitations, it ought to do so completely: no significant background should remain and the signal should be independent of bias intensity above some threshold. In the following, the "bleaching" bias illumination is obtained by passing the output from a 75W Oriel tungsten lamp through a silicon filter.

If the intent of the bias light is to create excitations, the resultant transient signal ought to be proportional to the density of such excitations; it may be possible to monitor this density with the dc transmission if the change induced by the bias is large enough. Note that a linear dependence on density does not necessarily mean that the signal should depend linearly on bias beam intensity; this depends on
the dynamics of the steady-state process. If the background "saturates" with bias, as for example the low-temperature "metastable" PA, any associated transient signals will do so as well. The "inducing" bias in the following is output from the same tungsten lamp passed through two IR-blocking filters and a 7800 Å long-pass filter.

**Figure 13.1.** Order of Survey presented in this chapter. We will examine spectra along the various one-dimensional slices of the $\omega_x - T$ plane in the indicated sections.

We now begin our overview of the TPA spectra obtained for various combinations of temperature and excitation energy. The order of our presentation is indicated in Fig. 13.1: we first examine the (weak)
\( \bar{\omega}_x \)-dependence at room temperature, and then the temperature dependence for high and low excitation energies, and finally the \( \bar{\omega}_x \)-dependence at low temperatures.

13.3 Excitation-energy Dependence at Room Temperature.

Figure 13.2 shows the TPA spectra obtained at room temperature for several different excitation energies (no intentional bias). The most striking thing about the various spectra is their similarity; they all show a monotonic rise by a factor of roughly ten between \( \bar{\omega}_p = 0.4 \) eV and 1.5 eV. All the spectra are thus difficult to explain in both their shape and their magnitude, as discussed in Chapter 6. Indeed, we showed data for 1.4 eV in that chapter (on a linear rather than semilogarithmic plot) as an indication of the MT-related thermalization, whereas the original observation had been made with much higher excitation energy. The similarity of the spectra validates that presentation.

The primary effect of excitation energy is on the time dependence of the overall magnitude: the amount of decay in one decade is indicated by the arrows. Although the shift of the spectrum (difference in length of arrows at high and low \( \bar{\omega}_p \)) is similar for all \( \bar{\omega}_x \), the high-\( \bar{\omega}_x \) data decay much more rapidly overall. This is consistent with the experiments of OKV, in which the shorter penetration depth and consequently higher excitation density cause bimolecular recombination, eventually decaying as a power law roughly \( t^{-\frac{1}{2}} \). For \( \bar{\omega}_x = 1.4 \) eV, the decay is negligible, as it should be for these long penetration depths. Indeed, I have checked that the pulse energy at which recombination begins for excitation with various penetration depths is consistent with the predictions of the MT model. Figure 13.3 shows an example.
Figure 13.2. Excitation-energy dependence at T=295K.
Figure 13.3. Effect of pulse energy $E^*$ on time dependence of TPA, showing the effects of recombination.

13.4 Temperature Dependence for High Excitation Energy

In contrast to the excitation energy, the temperature has an important effect on the spectral shape, as shown by Orenstein (1981). Figure 13.4 shows his data, taken with high excitation energies (either 1.9eV or 2.1eV — I don't know). These data are shown because they include
Figure 13.4. TPA spectra for high $\hbar \omega_X$ at various temperatures.
T=400K, which I have not repeated. The general trend is that the data get steeper on the semilogarithmic plot as the temperature is raised; at 400K the decay is almost exponential, rather than linear as it is at 295K. Unfortunately, because he normalized all the data at 1.4eV, the absolute time dependence as well as the magnitudes have been lost.

If one normalizes the spectra at various temperatures so that they are equal at \( \hbar \omega_p = 1.4 \text{eV} \) (Fig. 13.5), the results seem to indicate an increasing spectral shift at higher temperatures. It must be stressed that this procedure is arbitrary, however; one has no way to verify its appropriateness for different temperatures, in contrast to the single-temperature case in which the recombination provided such a check. Indeed, the simple model for the TPA would predict that the spectra should be the same for different temperatures, as long as \( kT \ln(\nu_o t) \) was the same; this is definitely not the case, as Orenstein demonstrated.

Since the spectrum is changing with temperature, it is impossible to proceed further with detailed modeling of its shape. One expects some broadening of any electron-phonon coupled lines, but not the dramatic changes seen in Fig. 13.4. It is therefore unclear how much faith to place in the apparent increase in thermalization rate: if the spectra cannot be overlaid by a simple shift of energies, how can one define the shift? In general, it is impossible.

The problem is dramatized in Fig. 13.6, in which the data are normalized on the low-energy side. Rather than suggesting an increasing thermalization rate, this figure suggests that the spectrum is the sum of two parts, and that the higher-\( \hbar \omega_p \) part becomes relatively more important at high temperatures. This possibility could help explain why the spectral shapes conflict with theoretical expectations (see Chapter
Figure 13.5. Data of Figure 13.4, normalized at $\hbar \omega_p = 1.4 eV$.

Indeed, the 20K data (which may represent a pure component) are similar to the theoretical ionization spectra. It also provides an easy explanation of the change of the spectrum with temperature.

Of course, the simple linear spectrum observed at room temperature can be no more than a coincidence, if this model is correct; indeed, the data at other temperatures are not linear in $\hbar \omega_p$. Since the thresholds of the linear fit provided the first dramatic evidence for the MT model, however, this would be a strange turn of events. It does seem likely that there are several components to the spectra, however. The present chapter describes an effort to sort out these contributions. While the effort has not been entirely successful, several interesting results
Figure 13.6. Data of Figure 13.4, normalized at low $\hbar \omega_p$. 

have been discovered.

13.5 Temperature Dependence for $\hbar \omega_x = 1.4 \text{eV}$

For high excitation energies, lowering the temperature increases the induced absorption at low probe energies relative to that at higher probe energies. This effect is also observed for low excitation energies ($1.4 \text{eV}$ in Fig. 13.7), except that the effect is even more dramatic: one observes a peak in the TPA, at around 0.6eV. The figure shows this peak continuously growing out of the monotonically increasing spectrum seen at room temperature for all excitation energies.
Figure 13.7. Temperature Dependence of TPA Spectra for $\bar{\hbar}\omega_x = 1.4$ eV.
The data shown in Fig. 13.7 are taken with no intentional bias. In fact, it was later discovered that the pump beam itself has a substantial effect on the data. In Figure 13.8 we show the normalized response to three trains of pulses, with the second train differing in pulse energy (and thus average power) by a factor of ten. The high pulse energy produces a larger cross section per excitation photon than the initial lower pulse energy: the signal is superlinear. This superlinear must correspond to an effect of the background associated with the train of pulses, because after the exposure, the low pulse energy produces even more signal per photon (sublinear intensity dependence); it must be re-exciting long-lived excitations created by the second train.

All of this has important effects on our understanding of the phenomena involved, as we will discuss. It also causes important difficulties with the reproducibility of the measurements; the two low-pulse-energy measurements differ by more than a factor of three! The introduction of bias beams helps to clarify the situation by (ideally) forcing the sample into a well-defined state. The effect of a "bleaching" bias ($\bar{\nu}\omega < 1.2 eV$) on the TPA signal is shown in Fig. 13.9. Clearly most of the observed signal is removed by the bleaching; presumably the "unbiased" signal results from re-excitation of the large background of absorption built up by the train of excitation pulses.

Note that the addition of an "inducing" bias ($\bar{\nu}\omega > 1.7 eV$) does not uniformly bias the sample into a saturated state (see Appendix C) since the high-energy light is absorbed in a thin surface layer. Thus the 1.4 eV light is free to induce more excitations throughout its much longer penetration depth. The observation of an increase when the "inducing" bias is applied implies that the density of excitations the
bias is creating is orders of magnitude larger than that created by the pulsed excitation.

At first sight the observation of a peak in the TPA is very exciting, since its narrowness appears to imply a localized-to-localized transition in a photo-excitation. The observation that it also requires a background makes this model much less plausible. Fig. 13.10 illus-
Figure 13.9. Effect of bias beams on the TPA signal for $\hbar \omega_X = 1.4\text{eV}$.

As the necessary model. By the time it has been excited three times, the excitation has absorbed roughly 3.5eV of energy from photons; that such an excitation should still not have enough energy to be extended, even with the necessary energy relaxation, seems unlikely. Furthermore, the presence of a significant long-lived induced absorption created by the bias suggests an alternate mechanism, one based on transient photo-
Figure 13.10. Simple but implausible mechanism for a peak in the TPA spectrum. Excitations created by the bias beam ($\hbar\omega_b$) are re-excited by the excitation pulse ($\hbar\omega_x$). The new excitations can be re-re-excited by the probe beam ($\hbar\omega_p$).

bleaching, or (TPB)

13.6 Transient Photo-bleaching (TPB)

We discussed in Section 12.3 the bleaching (reduction) of the long-lived absorption by light within the absorption band. Such bleaching should be observable in transient experiments as well. The results of an experiment using an inducing bias beam and a low-energy
Figure 13.11. Transient Photobleaching. The TPA signal resulting from low-energy pulsed light in the presence of an inducing bias shows induced absorption at low $\bar{\nu}\omega_p$, bleaching at high $\bar{\nu}\omega_p$.

($\bar{\nu}\omega_x=1.175$eV) pulsed excitation are shown in Fig. 13.11. The negative change in cross section for $\bar{\nu}\omega_p>1.2$eV shows that transient photo-bleaching (TPB) has been observed. Interestingly, however, bleaching is not seen in the whole induced absorption band: at low probe energies the absorption is enhanced, showing a narrow peak. The similarities between this peak and that observed for 1.4eV excitation suggest that for the
original data, the 1.4eV photons provide their own inducing bias.

Figure 13.12. Probable mechanism for photobleaching and peak in absorption. The re-excitation of the background carriers yields excess carriers at shallow energies and a relative deficit at deeper energies.

The most probable mechanism for this observation is illustrated in Fig. 13.12. The bias beam has created a high density of excitations, as manifested in the various long-lived processes such as the "metastable" induced absorption. Excitation within this induced band temporarily reduces the density of these excitations, which results in bleaching.
Some of the re-excited carriers will be found in higher-lying states, and these can give a greater absorption at low energies than would otherwise be possible.

In this model, the peak corresponds to the difference of two re-excitation spectra: a PA spectrum corresponding to absorption by the excited carriers, and a (negative) PB spectrum corresponding to the lack of absorption since those carriers are no longer in deeper states. It is surprising that these spectra cancel so precisely: the magnitude of the PB is only 1/10th of the PA at the peak, so the magnitudes must agree to within 10%. This may be coincidence. For $\hbar \omega_x = 1.4eV$, the net absorption does not become negative at any photon energy.

Evidence that the high-$\hbar \omega_p$ part of the spectrum results from a competition between different components is shown in Fig. 13.13. Depending on the pulse energy, the absorption can be either increased or decreased. Indeed, for the lower pulse energy, the signal appears to change sign with time delay, suggesting that the outcome of the competition changes as the distribution of excitation evolves.

It is clear that since the absorption coefficient of the virgin sample is very small at the probe energies used, the bleaching must involve removal of the extra absorption induced by the bias. Further evidence for this is shown in Fig. 13.14, where the dependence of the bleaching signal on probe intensity is shown. Recall that for a signal to be interpreted in a simple way, the result ought to be independent of the probe intensity; the results thus show that this condition is not satisfied, and that the probe must be treated as a bias beam. This is reasonable: at low temperatures, mechanisms that destroy excitations by thermal excitation are very slow. Optical bleaching can easily be the
dominant bleaching process, and this is the role that the probe serves. If one wishes to avoid extraordinarily low probe intensities, it makes more sense to introduce an inducing bias and a strong bleaching bias, and to control the level of excitation by their competition.

Such a competition was studied by Orenstein and Kastner (Kastner, 1980). As shown in Fig. 13.15, they studied the dependence of the steady-state absorption on the ratio of the intensities of a low- and a
Figure 13.14. Dependence of TPB signal (cross sections are negative) on the probe intensity.

high-energy beam. Although the data is somewhat noisy, it appears that the resulting absorption decreases as the square root of the ratio, above a certain threshold. This important result suggests that the bleaching process is bimolecular, probably a photo-assisted recombination of excitations. Returning to Fig. 13.14, we see that the transient photo-induced bleaching signal has roughly the same square-root depen-
Figure 13.15. Steady-state absorption resulting from competing "inducing" (intensity $I_h$) and "bleaching" (intensity $I_l$) beams.

dence on probe intensity. Thus, although this dependence is at first surprising, it provides a strong indication that the bleaching signal arises from the steady-state induced absorption produced by the competition of the intentional inducing bias and the probe beam.

13.7 Defects or Tail States?

In representing the transient bleaching effect in Fig. 13.12, we assumed that the long-lived absorption comes from a large occupation of
tail states. This is in marked contrast to the more customary description of the long-lived absorption as reflecting a saturated density of defect states, which constitute a discrete set of states in the gap. We have already noted, in the last two chapters, that the two main features that led to the defect description, the "metastability" and the "saturability", are on weak experimental ground.

In contrast, the multiple-trapping model and the associated thermalization have strong support; the hopping-down process discussed earlier is less well developed, but seems to be implicit in the band-tail description. If the long-lived absorption arises from band-tail states, then the occupation of those states presumably results from a balance between optical filling and thermal or optical emptying; at the lowest temperatures thermal emptying may be negligible. If there is only one beam on the sample, the eventual saturated density results from a competition between the inducing and self-bleaching tendencies of that beam. The density of centers may then be much higher than the $\sim 10^{17}\text{cm}^{-3}$ estimated by Bishop, Strom, and Taylor.

The saturation question should be susceptible to examination using transient techniques: If the states are truly saturated, then any new carriers excited cannot populate any more of them. The observed transient spectrum in the saturated regime should therefore differ from that of the "metastable" induced absorption (which is distinctively flat). Unfortunately, this determination is complicated by the fact that excited carriers can still absorb light, with unknown spectra, when they are not in "defects." There may also be a significant contribution from carriers re-excited out of the long-lived states. Nonetheless, although careful studies of the spectra as well as the time dependence must be
performed, the transient measurements in the presence of saturation may well hold the answer to this important question.

Further information may be gleaned by studying the dependence of the observed spectra on the photon energy of the pulsed (bleaching) excitation as well as the bias conditions and the evolution of the spectra with time. A face-value examination of Fig. 13.12 suggests that these quantities should play a direct role in the distribution of carriers and thus the resulting spectra, and they are all under experimental control. Fig. 13.14 suggests that the pulse energy is also an important factor (if this is not a bias-light dependence in disguise). A detailed study of the bleaching/peak region of the parameter space (low temperatures, inducing bias, and low excitation energy) should provide a wealth of information about the nature of the long-lived induced absorption and about the process by which it is bleached. Such information would almost certainly clarify the relationship between the defects and the band-tail states. This in turn might provide the basis for a true unifications of the diverse electrical and optical phenomena observed in chalcogenide glasses.

13.8 Excitation Energy Dependence at Low Temperatures

We complete our survey of the $T-\bar{\hbar}\omega_x$ plane with a discussion of the trends with excitation energy at low temperature, shown in Fig. 13.16. We have discussed at length the peaked spectrum produced by low excitation energies; clearly this is a complicated phenomenon and requires further experiments for complete elucidation. As $\bar{\hbar}\omega_x$ is raised, the spectrum evolves toward a monotonically increasing- or at least flat-
Figure 13.16. Excitation energy dependence of spectra at 22K.
spectrum; this evolution is reminiscent of the evolution of the \( \text{low-}\hbar \omega_\chi \) spectra with temperature. Probably there is no good reason for the similarity, which just reflects a change in the relative magnitudes of different components.

What the spectra fail to accentuate is that a very unusual process occurs in the middle of the illustrated range—near 1.7eV. This process is related to the luminescence (whose efficiency peaks near 1.7eV), and was in fact predicted (and then observed) by Robins and Kastner (1984) following studies of the luminescence of crystalline \( \text{As}_2\text{Se}_3 \). Those studies show a clear correlation in both time and temperature dependence between the luminescence and an induced absorption process. They associated the induced absorption process with the excitation labeled "PA" in Fig. 10.4.

In the amorphous material the situation is less clear, because there is an additional absorption process traditionally associated with individual free carriers. Nonetheless, as shown in Fig. 13.17, the time decay for the 1.7eV data shows a clear sigmoidal dependence. If one associates this dependence with a fast and a slow part, the fast part (region of downward curvature) shows the same correlation in time and temperature with the luminescence as found in the crystal; it is undoubtedly a related process.

Interestingly, the precise shape of the sigmoidal curve is sensitive to the bias conditions. Fig. 13.18 illustrates this in the dependence of the observed time decay on excitation pulse energy \( E^* \). The early-time part of the PA (associated with the PL) is not affected, while the later part (associated with free carriers) is larger for more energetic pulses. Presumably the latter increase reflects a contribution
Figure 13.17. Time dependencies at various excitation energies, for $\hbar\omega_p = 0.6$eV. Similar shapes are observed for $\hbar\omega_p = 1.2$eV.

to the PA from re-excited carriers; the bleaching bias is apparently not sufficiently intense. The constancy of the PL-related PA is consistent with the idea that one must create a strongly correlated electron-hole pair (exciton) if one is to observe luminescence.

Unfortunately, when the work of Robins and Kastner was done, the
The important role of the (unintentional) bias conditions on the results was not appreciated; such an appreciation might lead to clearer experimental differentiation between the PL-related PA and the other PA. It is surprising that the spectra measured by Robins and Kastner are the same for both the short-time and the long-time component. Such an observation, in which the exciton and free carrier give the same absorption, virtually demands an explanation in terms of one carrier weakly bound to a strongly-trapped carrier as was described for other reasons in chapter...
Returning to Fig. 13.18, it is intriguing that the transition between a constant induced absorption and a slow power-law decay occurs in the same region of $\tilde{\omega}_x$ in which the PL-related PA is observed. The latter process tends to obscure the transition, since the 100μs time chosen lies well within its sphere of influence (at longer times the results should be more clear). Nonetheless, it is interesting to speculate about a possible correlation between the two phenomena.

Figure 13.19 Time Decays for different pulse energies at $\tilde{\omega}_x=1.9$eV.
The time dependence at high $\tilde{R}_\omega$ itself is quite interesting. According to the multiple-trapping model, there would be no recombination at these temperatures; any recombination must involve a direct tunneling to recombination centers, as has been discussed by Pfost and Tauc (1983). This process is somewhat different from the tunneling thermalization that was a primary subject of this thesis, but the latter may affect the former. Figure 13.19 shows data for different pulse energies in the high-$\tilde{R}_\omega$ regime. Clearly the signal grows only sublinearly (cross section per photon decreases). This suggests a saturation or recombination phenomenon. What is intriguing about these data is that there is no evidence of a change in the time dependence; apparently most of the decay of the magnitude has occurred at earlier times. Although no attempt has been made to model this phenomenon, it seems to be an interesting one from the point of view of tunneling recombination.
APPENDIX A

DYE LASER TRICKS

This section describes a few helpful hints, discovered by trial and (expensive) error, for dealing with the Quanta-Ray YAG pumped dye laser. It should not be regarded as a replacement for, but as a supplement to, the Quanta-Ray manual. In addition to emphasizing a few topics not adequately covered by the manuals (for example, retro-reflections), it describes the function of the various adjustment in a way which I have found helpful in remembering and understanding the procedures for adjustment.

A.1 General Discussion

The proper adjustment of the dye laser requires the correct setting of some two dozen adjustments. If one considers this as a problem in finding the maximum output in a 24 dimensional space, the difficulties become obvious. The purpose of this appendix is to help reduce that task to a sequence of optimizations in a smaller dimensional space. Certain adjustments, however, are intimately connected with one another, and Fig. A.1 illustrates the importance of being aware of these connections. The figure shows a two parameter space, in which the response is optimum along a thin contour which is not aligned along either of the available adjustment axes. In this case, if one were to adjust the two parameters unknowingly, the optimization would follow a tortuous and time consuming path, as shown. By being aware of the connection between the parameters, the time can be significantly reduced, simply by overshooting the apparent optimum for each adjustment.
Figure A.1. Importance of principal axes. The system response is illustrated as a contour plot; the available adjustments are horizontal and vertical. The fastest optimization when the principal and adjustment axes are skew as shown is to overshoot the optimum with each adjustment.

The most useful thing to remember about adjusting the dye laser is that looking at the actual light output is often more useful than just monitoring the power output. There are several reasons for this. One is that there are several ways of achieving the same output power, so this measurement does not give any information about how further adjustments should be made. In contrast, the shape of the output beam will often give a direct indication of the way in which the adjustment is bad, for example if the spot is off to one side.

Another problem is that there may be several subsidiary maxima in
the power output; optimizing for power, one may think one is doing as well as possible, when in fact the output is poor. Comparing the power output at each stage with previously obtained values reduces the probability of problems, but is slow since it requires optimization before a comparison can be made.

The final reason for using visual inspection is that it is faster: power monitors typically have a response time of a few seconds (which is a very annoying time scale). In practice, this will not just make the time longer, but will make the eventual level poorer, because the user will become bored, and unwilling to try to get better performance. The recommended procedure is: Use visual inspection to get all adjustments close to their global optimum values, and only after all adjustments are close make a final optimization on the basis of the power output.

Unfortunately, the optimum position of some adjustments depends on how hard the laser is being pumped. This makes sense, since the a laser is an inherently nonlinear device, but it makes adjustments difficult. It is most reasonable to optimize the laser at maximum pump intensity, since non-optimal response at lower pump intensities can be compensated by strengthening the pump. Visual adjustments are poor at high intensities, however, because the eye's response becomes saturated and is incapable of discerning the true intensity distribution. A reasonable procedure, therefore, is to align the laser visually at low power, then proceed to high power and optimize the power response. Although this entails adjusting the laser with full pump intensity (and thus risking damage caused by misalignment), it seems to be the only way to insure truly optimal response.
A.2 Individual Adjustments

Various tricks to aid visual alignment are included in the following sections. The various adjustments to the laser are discussed in roughly the order in which they should be made.

A.3 Diffraction Grating

The single most common mistake made in aligning the dye laser is the failure to set the grating to the center of the dye response. Although trivial, this is a very serious mistake, since by the time it is discovered most of the other adjustments of the laser have often been misaligned in a misguided attempt to find the output. It may even happen that the laser will be aligned to optimize the amplified spontaneous emission (laser action not involving the grating). To avoid this problem, the operator should be aware of the approximate lamp energy (in J) which is necessary to barely cause lasing in each dye. If the response is not what it should be, the grating setting should be one of the first things checked.

A.4 Oscillator Stage

The oscillator dye cell is pumped by only a small fraction (around 10%) of the pump energy, and therefore operates much closer to threshold than the amplifier stage. The adjustments of the oscillator stage are therefore quite critical, and also, because of the small size of the excited region, quite sensitive. Furthermore, there are many (six) separate adjustments to be made, and there are several subsidiary maxima in the output. For all these reasons, it is important to be familiar with the effects of the various misalignments, both their theoretical role in the lasing process and their signatures in the output beam. The
signatures are best learned by direct observation of the output under various circumstances. In this section I will describe some heuristic arguments that make the six dimensional space seem a little less formidable.

Figure A.2. Beam shaping geometry. The output from the YAG laser and Harmonic Generator is a donut shape, roughly 1cm in diameter. The beam splitter generates a reflection from each surface, and the cylindrical lens focusses these almost to a line by the time they reach the oscillator dye cell.

For pumping with doubled or tripled Nd:YAG output, the pump beam entering the laser is roughly donut shaped. Fig. A.2 shows how this
donut is changed by the optics into the rather strange, roughly linear spot shown in Fig. A.3. The reflection from the front and back surfaces of the beamsplitter overlap somewhat, but are rather separate, due to the thickness of the beamsplitter. The direction at which these overlapping donuts leave the beamsplitter, and therefore the position of the pump spot on the dye cell, can be adjusted by means of allen-head screws on the mount of the beamsplitter. Note that if the angle or position of the incoming pump beam is changed (for example by adjusting the position or angle of the output turning prism in the prism harmonic separator), the tilt of the beamsplitter must be changed to compensate. For this reason, it should be confirmed that the pump beam is properly hitting the amplifier turning prism before adjusting the beamsplitter tilt. (The incoming beam must be adjusted when adding or removing the preamplifier beamsplitter.)

Before hitting the oscillator dye cell, the double donut from the beamsplitter passes through a cylindrical lens, which focuses it along one direction. The direction along which it is collapsed can be changed using a vertical 1/8" allen screw in the cylindrical lens mount. The degree to which the beam has been focused by the time it reaches the dye cell is determined by the position of the lens: the focus is a fixed distance from the lens, since the input beam is collimated. It is important to realize that the concentration of dye molecules in the oscillator mixture is sufficiently high that the absorption length of the pump light is less than a millimeter. The excited region of dye is thus a thin sheet, whose lateral characteristics are determined by the shape of the beam when it first encounters the dye. It is these characteristics that are altered by the various adjustment screws.
Figure A.3. Oscillator dye cell geometry (also applies to preamp, if used). The laser output is determined by the overlap between the focussed pump and a roughly linear region in which the emitted photons are reflected back again. Note the important interaction between angle and position (see Fig. A.1).

It is highly recommended that all the gross and medium-scale adjustments be made by looking at the light output of the oscillator directly. This is most easily done by placing an index card just in front of the second (positive curvature) lens of the telescope (see Fig. A.4). At this point the output beam has already been expanded by the first lens, so the various features are more easily observed.

The observed output will typically have several bright spots, which represent various pathways for amplification of the laser fluorescence. Only one of these spots represents the optimum pathway, and the purpose of the coarse adjustments is to insure that most of the light is in that spot. It should be clear from this description why an optimization purely on the basis of power output is such a haphazard procedure. Often
Figure A.4. Telescope geometry. In addition to identifying important retro-reflections, this figure shows the best locations to monitor the oscillator/preamp output as well as to check for retros.

the "good" pathway will only be bright over a very narrow range of adjustment, with a dramatic shift to other spots when the parameter is moved outside this range.

One type of subsidiary spot that deserves special attention is the amplified spontaneous emission (ASE) spot, which results from a reflection from the beam expanding prism or a similar non dispersive element. As a result it has a wavelength near the peak of the dye response, independent of the setting of the grating. This means that if the grating is set far from the peak, the ASE spot may be the only response present, but an optimization of this response would result in a truly abominable alignment of the "good" spot. One should be suspicious of this possibility (which can easily be checked by changing the grating and looking for a color or intensity change) if the laser seems to be
grossly out of alignment.

Now the various adjustments of the oscillator stage will be described. It is sometimes helpful to think of these adjustments in terms of their effect on the spot shape, as illustrated in Fig. A.2. The pump creates a region of the dye which has high gain for photons with an appropriate energy. Photons emitted in a certain region (which I call the amplification region) will have a large probability of being reflected back into the same region. This amplification region is roughly sketched by cross-hatching in Fig. A.3. The resulting output of the oscillator will depend on how much of the high gain region overlaps the amplification region. Because of the complex shape of the high gain region, there are many subsidiary maxima in the output power.

**Cylindrical Lens Angle.** The angle of the lens determines the inclination of the line of gain and should be roughly horizontal for proper lasing. It is important that this adjustment be relatively good before others, such as beamsplitter vertical, are optimized. Observing the fluorescence from the pumped part of the dye and adjusting this glowing line to appear horizontal to the eye seems to do an adequate job, for coarse alignment. It can then be optimized for power after the other adjustments have been made. Once it has been set, this adjustment should not need realignment.

**Beamsplitter Horizontal Tilt.** This adjustment is also done best by eye: the pump beam should be centered on the dye cell, so that both lobes of the beam are hitting dye. It should not be readjusted except for a final power tweak, since the eyeball alignment is quite satisfactory.

**Beamsplitter Vertical Tilt.** It is clear from Fig. A.2, and from
the long moment arm of the beamsplitter, that this is a very important
and sensitive adjustment. As the beam position is raised, the light out-
put will shift from being mostly below center, to near center, to above
center, in a fairly discontinuous pattern. The center position is the
desirable one, and by scanning through a large range of the screw the
optimum position should be obvious. The center spot is dominant only for
a fraction of a turn of the screw. It is important to familiarize one-
self with this critical adjustment before the laser gets badly out of
alignment.

Output Coupler Horizontal. This adjustment, a 1/8" allen screw
(also adjustable by hand), is rather poorly located at the bottom of the
plate holding the partial reflector which is the output coupler for the
laser oscillator. Presumably it changes the horizontal position of the
amplification region. Its effect on the output beam is similar to that
of the beamsplitter vertical, but in the horizontal direction. It is
also very sensitive. The vertical angle of the output coupler affects
the calibration of the grating, and should therefore not be changed.

Position of the Cylindrical Lens. The position of the lens deter-
mines the vertical extent of the pump beam at the dye cell. Since the
beam is more or less double humped in this direction, this can have a
rather complex effect on the output. For example, if the beam is not
centered on the amplification region, the maximum power might be
obtained by making the height such that one of the side lobes is cen-
tered. (A similar situation is described in the section on the amplifier
alignment.) It is thus recommended that the vertical position be optim-
ized first, with the focus one or two centimeters behind the front of
the dye cell, and only afterwards adjust the focus. This is a very
insensitive adjustment. **WARNING:** under normal operation, the focus of
the lens is at least one or two centimeters behind the front of the dye,
and the light is absorbed before it reaches the focus. If the focus gets
too close to the front of the cell, the intensity can become high enough
to burn the dye or the dye cell: this should be avoided.

Alignment on the basis of visual inspection of the output beam
should be sufficient to insure that the laser is being operated close to
the absolute maximum, rather than a secondary maximum. After all
adjustments have been cycled through (especially the beamsplitter verti-
cal and the output coupler horizontal) a final optimization can be made
by looking at the output power, although one should be careful not to
move any sensitive adjustment by a large amount. The easiest way to mon-
itor the power output is at the output of the laser, with the amplifier
beam blocked (for the present). Some of the light is absorbed in the
amplifier dye cell, but this is of little consequence.

A.4 Preamplifier Stage

The preamplifier geometry is similar to that of the oscillator, and
the descriptions in that section should suffice. The exception is the
dye cell translation, which replaces the output coupler horizontal as
the horizontal beam adjustment. Also, instead of an invisible region due
in which emitted light is returned to cause amplification, the important
region here is the region through which the oscillator output passes.
This region can be observed visually, if the pump beam is blocked.
Indeed, all of the adjustments of the preamp stage can be made quite
well by overlapping the pump beam with this region by eye. The dye cell
should be translated so that the oscillator beam passes just inside
front surface of the liquid. This condition can easily be determined looking at the output: two separate spots merge into one.

A.5 Beam Expanding Telescope

The telescope consists of two lenses: a negative curvature, diverging lens, and a positive curvature converging lens. This design allows telescope operation without only a virtual focus, avoiding the high power densities present at a real focus. There are two criteria for adjustment of the telescope: proper divergence properties of the output beam, and safe disposition of retro-reflections.

The divergence of the beam leaving the telescope determines the divergence of the amplified beam. If the dye laser output is used directly for excitation, the divergence can be set wherever desired, although if the beam is brought to a focus it is likely that it will cause breakdown of the air, or at least cause damage to any object brought close to the focus. If the dye laser is used to pump the Raman Shifter, however, extra care must be taken, because the Raman Shifter already has a focusing lens in it, in order to provide the high power densities necessary for stimulated Raman scattering. The focus of the pump beam normally occurs around ten centimeters from the far end of the pressure tube, if the dye laser output is collimated. However, if the dye laser output is made somewhat divergent, the focus will move near the output window of Raman Shifter, which can (and has) cause damage to that window, which costs several hundred dollars to replace. This causes degradation of optical quality, as well as a risk of an explosion if the output window were to break. **WARNING**: Do not adjust the divergence of the dye laser output beam while pumping the Raman Shifter. Do not operate the Raman Shifter with other than a collimated pump beam.
A.6 Retro-reflections

The most frustrating aspect of aligning the dye laser is the fact that a back reflection of only 5% of the output power can be reprocessed by the optics in such a way as to damage the laser itself. This possibility requires that alignment be performed in a very careful way, and no set of guidelines is a replacement for patience and care. This section should help, however, by describing the origins of the most important retros, and the adjustments that allow them to be disposed of safely. Although most of these alignments will have to be done only rarely, it is good for every user to be aware of the issues. The oscillator end pump needs to be adjusted quite often.

![Diagram of retro-reflections](image)

Figure A.5. Elimination of retros. Several degrees of freedom are available to send the retros away from the optical axis while keeping the forward beam on axis.

The first retro to be discussed is a reflection from the flat surface of the negative curvature lens. (The reflection from the curved surface is diverging, by design.) This reflection, if it were sent back
into the oscillator, could potentially be amplified in the dye, and possibly cause damage. Although this is unlikely, the negative curvature lens mount is normally rotated clockwise (looking down) as far as it goes (without removing the mounting screws) in order to send the reflection off to one side.

Before discussing the other retros, let us examine the operation of the telescope, shown in Fig. A.5. The (real and virtual) foci of the two lenses should be in the same plane, which is slightly before the negative curvature lens (that is, closer to the oscillator and preamp). This means that any backward-travelling collimated beam encountering the curved surface of the positive lens will be focused to the same plane, which is not very far from the negative lens. If the power in such a beam is large, it can cause damage to the negative lens, requiring careful adjustment of that lens to avoid degradation of the beam, or even replacement at significant cost. Different adjustments are used to insure that the various retros, numbered in Fig. A.5, do not cause this damage.

Retro 1 comes from the second, flat surface of the positive lens. (The reflection from the curved surface is divergent, by design.) In order to send this retro off to the side, and onto the heat absorbing glass attached to the negative lens’s mount, the positive lens is canted clockwise, in a similar fashion to the canting of the negative lens. This retro can be seen even when there is an index card placed between the positive lens and the turning prism, which distinguishes it from the rest of the retros.

Retros 2 and 3 come from the two flat surfaces of the turning prism. They can be recognized because they are still present when an
index card is placed after the prism, but not if it is placed between the lens and the prism. To send these back at an angle, one might think that it would be sufficient to simply change the angle of the prism (accessible through an access hole in the front panel). Unfortunately, this adjustment also changes the angle of the beam going to the amplifier cell, so another motion is necessary to compensate. This second motion is translation of the positive lens, in the direction perpendicular to the beam (away from the front wall). As seen in Fig. A.5, since the original beam hits the lens away from the center, the beam leaves the lens at an angle. It should be noted that the output of the lens will no longer pass through the alignment hole on the sidewall of the laser, when the turning prism is removed, but this is of no true importance. Under these conditions, adjusting the turning prism so that the output is parallel to the dye laser sidewall leaves the retros 2 and 3 to one side of the negative lens. The distance they move is approximately the distance by which the positive lens was translated. Translating the lens by a moderate amount should have little effect on the actual beam shape. [This section has described my own inferences about how to align the laser to avoid retros; I suspect that the Quanta-Ray people think about things somewhat differently, and some of the differences could be important.]

As a result of the concerted movement of the positive lens and the turning prism, the oscillator beam is turned by exactly the 90° desired, but the retros from the lens and the prism should be deposited on the heat absorbing glass, as shown in Fig. A.6. This means that if a retroreflection from somewhere past the turning prism is reflected in such a way as to travel exactly back along the beam path, it will be focused
Figure A.6. Proper disposition of retros and reverse travelling pump onto heat-absorbing glass.

very near the center of the negative curvature lens, with potential for damage. There is no adjustment of the telescope/turning prism assembly which will prevent this problem; any retros from further down the line must be taken care of at its source. Note that since these retros will be a fraction of the amplified laser output, they have much greater potential for damage than the retros discussed so far, especially since they may pass through the amplifier cell themselves.

A specific example of this sort of retro is the reflection from the Raman Shifter, generally from the input polarization rotator, or perhaps even from the input window of the pressure cell. This retro has in fact actually caused damage to the negative lens in our laser. It should be noted that to avoid these problems, one will often have to overcome a natural tendency to line everything up "square," since it is square alignments that cause retros to retrace their paths exactly.

The fifth source of reverse-travelling light in Fig. A.4 is actually the most important. It is not actually a retro-reflection, but is a fraction of the pump beam for the amplifier cell. If the amplifier is
being end-pumped, then the pump beam comes into the cell almost antiparallel to the output of the turning prism. As we have noted, this presents a danger of burning the negative curvature lens. This is especially important since the pump beam has an especially high power density. Fortunately, however, the small angle the pump beam makes with the telescope/turning prism output means that the pump light, which travels straight through the amplifier cell, is deposited safely: most of it hits the absorbing glass mounted on the positive lens mount, the remainder is focused and hits the absorbing glass on the right side of the negative lens. This beam can be seen as a donut shaped beam, and its progress through the optics is easily followed.

The danger comes from pump light that does not take this standard path. For example, if the amplifier cell is off-center in relation to the, pump some of the light will be piped through the edges of the cell. It will not be attenuated by passage through the dye, and it may come off at any angle. It is thus capable of causing significant damage to the negative lens, or to other parts of the laser, if conditions are wrong. Thus the adjustment of the amplifier dye cell position and the amplifier pump position must be done in tandem, so that the pump always hits the cell squarely. I recommend that the reader deliberately perform this misadjustment (at minimal power level!!!) to see this for effect himself or herself. It is normal for a small fraction the pump to be focused near the negative lens; at least, I have not been able to eliminate it.

All adjustments in this section should be performed at the lowest power level at which the light in question is visible. In this way the effects of gross maladjustment can be seen without danger of damage. The
proper way to look for retroreflections is to use an index or business card. Slide the card in slowly from all four directions, until the card blocks the main beam. Any spots that disappear when the main beam is blocked are retros. Of course it is possible for the retro to be exactly collinear with the main beam, in which case it will not be visible by this technique. For this reason, the user should try to find all the reflections listed here, and also look carefully for any additional reflections. It is also advisable, when a new configuration is being used for the first time, to search carefully for retros at several power levels between the minimum and the maximum, since some retros which are dangerous at high power levels may not be visible or seem important at lower power levels.

A.7 Amplifier Stage

The amplifier stage is the simplest one to align, in part because the beams are relatively large, and the adjustments are therefore not as sensitive. In addition, the output of the dye laser (although not of the Raman Shifter) depends only more-or-less linearly on the amplifier gain, rather than being strongly superlinear as for earlier stages. If the amplifier is sidepumped, the adjustments are very similar to those described for the preamp above, except that instead of the beamsplitter there is a turning prism, and instead of a cylindrical lens there is a cylindrical mirror.

For end pumping, however, there is, a certain ambiguity in the appropriate alignment, resulting from the fact that (for end pumping) the region of gain is roughly donut shaped, while the (expanded) oscillator output is more or less Gaussian. As shown in Fig. A.7, there are two reasonable ways of trying to align these regions. In the symmetric
Figure A.7. Alignment of Amplifier End-pump. Because the pump has a donut-shaped intensity distribution (shown here in cross section), while the incoming beam has only a single maximum, there are two ways to achieve substantial overlap.

alignment, the pump runs through the center of the donut. This will produce the most symmetrical output beam, but because the maximum in the oscillator beam runs through a region of low gain, it may or may not produce the maximum total power. This is the alignment preferred by the person who installed our dye laser. The second possible alignment abandons one side of the donut entirely, in order to place the maximum of the oscillator at a position of high gain. The output will be less symmetrical (although it can be surprisingly good), but this arrangement may produce more total output, especially with dyes that are close to threshold. At full power, however, the symmetrical arrangement appears to be the best. A further complication is the fact that, since the spatial distribution of power (the beam profile) is different in these two cases, the response of the Raman Shifter will not have the same dependence on the dye laser pulse energy, so that the arrangement which gives
more dye laser output may give less Raman output. This latter possibility also depends on the order of Raman shifting. These questions have not been resolved at the time of writing.

To grossly align the amplifier cell, it is easiest to observe the scattered light from the oscillator beam to ascertain its position, while the pump beam is blocked. The dye cell can then be moved into the desired position with respect to this beam, and the pump beam can be aligned (at low intensity!) to the proper position on the dye cell. The position of the pump is adjusted by screws on the turning prism, similar to those on the oscillator beam splitter. As discussed above, the adjustment of the horizontal position of the pump should always be accompanied by a motion of the dye cell, so that the pump remains centered on the cell, especially at high power levels. The position of the dye cell itself is otherwise unimportant. The output spot can be observed (through a color filter, if desired, to reduce the brightness) and the effects of aligning the two beams symmetrically or asymmetrically determined. However, if the Raman Shifter is being used, it may be necessary to perform final adjustments on the basis of the Raman output, which unfortunately requires full power operation.

A.8 Power Tweaking

Once all the laser adjustments are close to their optimum positions, it is a good idea to make a final adjustment, looking at the output of the laser on a power meter, for example the Scientech. This is necessary because the logarithmic response of the eye, and its saturation at the high intensities typical of the laser output, make it essentially impossible to see the fine distinctions in brightness. If the Raman Shifter is used, however, its nonlinear response will be very
sensitive to these small variations. It is also useful to measure the power output, in order that the laser output may be compared over extended periods as a warning of possible problems.

In making power tweaks, it is imperative that two quantities be known: the time response of the detection system, and the sensitivity of the adjustment being made. The time response must be known so that one has an appreciation for the meaning of the signal displayed. One should be careful not to respond to changes in the output any faster than the time constant of the system. Such overly quick response means that one is responding to a situation which no longer exists, and many problems, including oscillations, can arise. I have found it very difficult to respond in a continuous fashion at a slower rate than about a second. Since most power meters have a slower response time than this, a new technique is necessary. The technique I use is to move in discrete steps, of the characteristic size appropriate to the adjustment being made (see below). I then wait for the meter to respond to the new situation before making another step. In this way, I can force myself to respond more slowly than the natural human time scale.

The importance of knowing the sensitivity of the adjustment is illustrated in Fig. A.8. The sensitivity determines the characteristic size of a step required to make a perceptible change in the response. This change will not show up on the meter for a time comparable to the system response time. If one makes the adjustment too fast, moving several characteristic sizes in a response time, then the peak in the response will be washed out (as well as displaced), and it will be very difficult to find the peak. If one is close to the peak and moves the adjustment by many characteristic sizes, it will take a very long time
Figure A.8. Effects of power monitor time constant. If the sweep rate is such that the entire peak is crossed in less than one time constant, the peak of the monitor output will be displaced from the true peak, and will also be less pronounced.

to get back where one started, a truly frustrating experience. If two such mistakes are made, one might as well give up, and go back to visual inspection. In adjusting the dye laser for maximum power, be sure that the pump power is not drifting (for example, due to heating of the doubling crystal) by going back and forth across the peak a couple of times to be sure it's real.

A.9 Raman Shifter

Up to this point we have said very little about the Raman shifter, except that reflections from it can damage the dye laser. The Raman shifter provides the ability to tune the output photon energy over an
incredibly large photon energy range. It is also simple to operate, being simply a tube of pressurized hydrogen.

The only parameter that really makes any difference to the output is the pressure of Hydrogen in the pressure cell. In principle the optimum pressure may vary depending on the order of Raman shifting desired, but I usually operate with about 300psi for all orders. I have found that it sometimes helps to refill the tank with fresh hydrogen after a few weeks, however. Presumably the gas becomes contaminated after a while. This is worth trying if things aren't working right.

**Safety.** I usually vent the Hydrogen directly into the room. The reasoning is that the pressure cell and associated tubes have a volume of about 1 liter; if they are at 300psi (20 atmospheres) this corresponds to 20 liters of Hydrogen. To get self sustaining combustion (pronounced "explosion"), there must be at least 4% hydrogen. Thus once the Hydrogen is diluted to 500 liters, it's safe. Since 500 liters (ten cubic feet) is much smaller than even a small room, it seems fairly safe. This calculation is only reasonable if the Hydrogen is really mixed with the air, however: on its way out it can have significantly higher concentration. Please be careful, and let the gas out slowly.

If for any reason the angle at which the dye laser output comes into the Raman shifter is changed, the beam should be put back on track with internal Raman shifter adjustments. The input turning prism assembly (which also rotates the polarization) has a vertical and horizontal adjustment screw. These should be adjusted so that the beam finds the turning prism on the other side of the pressure cell. At present these are adjusted in a rather skewed fashion, so that the reflection from the pressure cell does not get back through the dye laser telescope and
cause damage (see Section A.6). Once the first angle is adjusted, the vertical adjustment on the turning prism should be adjusted so the beam makes it out the output slits. The horizontal adjustment is redundant with the micrometer drive. In making the vertical adjustment, be careful that the beam is not directed too far from horizontal, since there will then be danger of eye damage. In addition, it is possible to cause the beam to reflect off of the beveled surface of the pressure gauge in a highly dangerous fashion. As always, perform the alignment at minimum power.

In the discussion of the focusing of the dye laser telescope, we noted that an improper adjustment causes the output window of the Raman shifter pressure cell to develop damage; this seems to happen even when the telescope is adjusted properly. It may be necessary to adjust the prism angles discussed in the previous paragraph to avoid minor burn holes. Watch out for extreme damage, however, since if the (7mm thick) window loses its strength, there is a definite explosion hazard.

Note that the divergence of the Raman shifter output can be adjusted with the lens position near the output of the pressure cell.

The Raman Shifter operates by stimulated Raman scattering, and therefore needs a certain threshold pulse energy to operate. This threshold grows with the required order of scattering, but a good estimate is 30mJ out of the dye laser per pulse. Although the good dyes have no trouble supplying this energy, the marginal dyes are much harder to use, since the dye laser and YAG laser must be very well adjusted to get Raman shifting action. This should be kept in mind when choosing dyes for Raman operation.
APPENDIX B
CURRENT MEASUREMENTS

The accurate measurement of transient current involves some subtleties which although not difficult, require some simple modeling of the system before the problem can be successfully tackled. However, because they involve the nitty-gritty details of experiments, such as field and illumination uniformity of sample geometry, they tend to be omitted in books or articles that focus on the more fundamental aspects of the problems. This is an attempt to fill some of these gaps.

B.1 Space Charge and Field Distribution

Before discussing transient measurements, we will examine dc measurements, both with and without light. Some of the discussion here will also apply to the transient measurements, but more importantly, the dc conditions are the background against which the transient currents are performed and measured, and may for example, alter the field distribution seen by the transient carriers.

We begin by conceptually dividing the solid into small volume elements—elements small enough that fields, current densities and the like can be considered constant throughout the element, but large enough that the carriers in each element are in equilibrium with one another. (It can happen that the sample is so small that this division is impossible; this is the "hot electron" regime.) Then we can characterize each element by a local conductivity tensor $\bar{\sigma}$, which we will take to be a scalar $\sigma$, so the current density is

$$\mathbf{J}(\mathbf{r}) = \sigma(\mathbf{r}) \mathbf{E}(\mathbf{r})$$  \hspace{1cm} (B.1)
The rate of change of charge density $\rho$ in any element is just the amount of current entering it:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}$$

$$= -\sigma \nabla \mathbf{E} - \mathbf{E} \cdot \nabla \mathbf{v}$$

but $\nabla \cdot \mathbf{E} = 4\pi \rho / \epsilon$, so

$$\frac{\partial \rho(\mathbf{r})}{\partial t} = -\frac{4\pi \sigma}{\epsilon} \rho(\mathbf{r}) - \mathbf{E} \cdot \nabla \mathbf{v}$$ \hspace{1cm} (B.3)

In steady state there is no net change in charge density ($\partial \rho / \partial t = 0$). If the conductivity is uniform ($\nabla \sigma = 0$), then Eq. (B.3) shows that there can be no net charge density: the material is neutral in regions of uniform conductivity. If we begin with a situation in which there is a net charge density, it will be neutralized exponentially in time, the time constant being the Maxwell dielectric relaxation time,

$$\tau_M = \frac{\epsilon}{4\pi \sigma}$$ \hspace{1cm} (B.4)

The reader may find it amusing to reproduce this result by calculating the RC time constant of a parallel plate capacitor filled with a material of conductivity $\sigma$ and dielectric constant $\epsilon$. For highly insulating materials, the dielectric relaxation time is quite long. For example, with $\epsilon = 10$ and $\sigma = 10^{-12} \Omega^{-1} \cdot \text{cm}^{-1}$, $\tau_M \approx 100 \text{s}$.

These simple results apply when the conductivity is independent of space and time. In general the conductivity of a semiconductor depends on the electric potential (because of the change in the position of the Fermi Level with respect to the bands) as well as the time history of the potential (because of filling and emptying of traps), so that the general validity of these discussions depends on the microscopic nature of the system. In particular, the density of free carriers is often
quite small in amorphous materials and may be insufficient to screen a reasonable applied field. Screening is then limited by release from traps, or even bulk generation, and may exhibit anomalous time dependence reflecting the exponential DOS (Kastner, 1983). However, the dielectric relaxation time is a useful characterization of the small signal response of any material to an excess charge distribution.

![Diagram](image)

**Figure B.1.** Build-up of space-charge resulting from inhomogeneous conductivity.

If the conductivity is inhomogeneous, Eq. (B.3) shows that the steady state will be characterized by an accumulation of charge in regions of changing conductivity:

$$\rho_{\text{steady-state}} = \overline{E} \cdot \nabla \varphi$$  \hspace{1cm} (B.5)

the origin of this buildup is easily seen by studying the simple example in Fig. B.1, in which there is a single discontinuity in the conductivity at a plane perpendicular to the electric field. The current den-
sity in both parts of the sample must be equal, but since the conductivities are different, the electric field must be higher in the lower conductivity region. This arises from a buildup of charge at the discontinuity because of the initially higher current density when the fields were equal. The time taken for the sample to reach steady state is the dielectric relaxation time of the more conductive material. This is exactly analogous to a problem of resistors in series, in which the voltage falls across the higher resistance.

This effect of inhomogeneities in the conductivity can be important in steady state photoconductivity measurements, since nonuniform illumination will cause the field across the more conductive parts to be reduced. If the photoconductivity is higher than the dark conductivity, then uniformity of the illumination is essential to correct measurements. In transient measurements, as we shall see, this is less likely to be a problem. Still, if the average photoconductivity due to the transient illumination is larger than the dark conductivity, the magnitude of the current will be reduced in a similar way.

B.2 Transient Currents

Having discussed charge relaxation and the eventual steady state, we can now define transient currents unambiguously: transient currents are currents that are changing faster than the relevant charge relaxation time (which may or may not be given by Eq. (B.4)). The measurement of such transient currents sometimes confusing because what is actually measured is a displacement current, This concept is usually poorly understood, so I will review it briefly.

One of Maxwell's equations is (in Gaussian units)

\[ \nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c} \mathbf{J} \]  (B.6)
where $\mathbf{J}$ is the current density associated with motion of charges. The term $1/c \, \partial \mathbf{D} / \partial t$ is necessary to maintain the symmetry between electric and magnetic fields. It is also necessary for consistency with charge conservation, as we now show. Taking the divergence of both sides, and recalling that the divergence of a curl is zero, we get

$$
\nabla \cdot (\nabla \times \mathbf{H}) = 0 = \frac{1}{c} \frac{\partial}{\partial t} \nabla \cdot \mathbf{D} + \frac{4\pi}{c} \nabla \cdot \mathbf{J} \tag{B.7}
$$

But $\nabla \cdot \mathbf{D} = 4\pi \rho$, so we have derived conservation of charge from two of Maxwell's equations. Thus the term containing the time derivative of the electric field must be present in order that Gauss's law remain valid when the charge moves. The quantity $(1/4\pi) \partial \mathbf{D} / \partial t$ is called the displacement current.

This means that the quantity in brackets in Eq. (B.7) has zero divergence, for any fixed time. That vector field can thus be represented by closed loops (just like the magnetic field) in which the density of lines per unit area represents the magnitude of the "current." Associated with any real current $\mathbf{J}$, there must be closed loops of which that current density is a part. For example, for a slowly moving point charge in free space, the current loops have the same form as the field from a dipole, as shown in Fig. B.2(a). (We are ignoring relativistic effects—radiation—here.)

Now consider the same charge, moving with the same velocity, placed between a pair of capacitor plates, as shown in Fig B.2(b). On time scales which are long compared to the response time of the circuit to which the plates are connected, charge will flow from the circuit into the plates so that the field inside the electrodes is zero. The surface
Figure B.2 Displacement Current. (a) For a charge moving in free space, the displacement current has the form of dipole fields, that is, the difference in fields between two charges next to one another. (b) For a charge moving in the vicinity of electrodes, motion of charge in the electrode should also be taken into account. For the case shown, the circuit response is assumed fast, and charge is able to move from one electrode to the other to screen the field. Only the field lines from the charge that intersect the plates causes measurable current.

charge density necessary to screen the field in this way is given by 
\[ s = (1/4\pi)D_n \] , where \( D_n \) is the component of the field perpendicular to the surface.

The motion of the charge creates a change in the electric field at the surface, and therefore in the amount of surface charge necessary to screen that field. The amount of current which must flow into a unit
area to keep the field screened is just \( ds/dt = (1/4\pi) \partial \mathbf{D}/\partial t \), that is, the amount of displacement current crossing that area. This is the microscopic process by which the displacement current associated with moving charge is converted into actual charge flow in the external circuit. This is what we are actually doing when we measure a transient current.

The figure makes clear how a moving charge causes a current in an external circuit. The response of the circuit is complicated, however, because the electric field is the superposition of the field due to the charge directly with the induced charges on the electrodes. For infinite parallel plates, we could solve this problem using the image charge method, but for more complicated geometries this is not possible. Fortunately, a simple gedanken experiment allows us to relate the induced current to more familiar properties of the geometry. Consider the situation illustrated in Fig. B.3, in which we apply a test voltage \( V_t \) to a pair of electrodes of arbitrary geometry.

![Diagram](image)

**Figure B.3.** Determining the effect of electrode geometry on the measured current.

Examining a small volume element \( d^3r \) at \( \mathbf{r} \), the amount of power dissi-
Examining a small volume element $d^3r$ at $\vec{r}$, the amount of power dissipated in that volume is just $\mathbf{j}(\vec{r}) \cdot \vec{E}_t(\vec{r}) d^3r$, where $\vec{E}_t(\vec{r})$ is the electric field resulting from the test voltage $V_t$. This dissipated power (or generated power, depending on the relative direction of $\vec{E}_t$ and $\mathbf{j}$) must be drawn from (or delivered to) the battery. By setting the total power dissipated power at all positions $\vec{r}$ equal to the power delivered by the battery, which is $iV_t$, we get

$$i = \left[ \int d^3r \mathbf{j}(\vec{r}) \cdot \frac{\vec{E}_t(\vec{r})}{V_t} \right]$$

Thus the factor $\vec{E}_t(\vec{r})/V_t$ represents the contribution of a unit of current density $\mathbf{j}$ at the position $\vec{r}$, which is the factor we were looking for.

For example, in the sandwich geometry of Fig. B.2(b), $\vec{E}_t/V_t$ is simply $1/L$, where $L$ is the plate separation. Thus for a given number of carriers moving in the same electric field, the current in an external circuit is much larger for smaller plate spacings. Of course, if the excitation density is the same, then the number of carriers is proportional to $L$, and the thickness doesn't matter, for constant field. This is the reason that this geometrical factor may have escaped your attention before.

Note that the application of the voltage $V_t$ was only a mental exercise to derive the factor $\vec{E}_t/V_t$. This factor is a property of the geometry, and the current derived will show up in the external circuit whether or not a voltage is actually applied, for any given current density $\mathbf{j}(\vec{r})$. For example, the current $\mathbf{j}(\vec{r})$ in Eq. (B.7) need not be a response to the electric field applied to the electrodes; it may be a response to internal electric fields (for example, in a solar cell) or
diffusion of carriers. The response of the circuit to a given \( \vec{j}(\vec{r}) \) will be the same.

In performing the thought experiment to determine the factor \( E_t/V_t \), what other charges should be allowed to move? The answer is: any charges that are able to respond on a time scale shorter than the time scale for which the transient current is to be measured. For example, the reason the electrodes caused a complication was that the charges within them were able to be removed or supplied by the external circuit on the important time scale. On the other hand, although applying a voltage to a Schottky barrier or other space charge region will change the distribution of charges within that region (and therefore the electric field at the electrodes), this effect should be included only if it occurs fast enough to exert influence on the time scale of the experiment.
APPENDIX C

ISSUES IN GEOMETRY

The measurements described in this thesis require careful attention to the geometry of the various light beams. Some of the important issues are discussed here.

C.1 Photocurrent Geometry: Inhomogeneous Fields

Except for the infinite-parallel-plate geometry, no electrode configuration gives uniform electric fields everywhere. The coplanar geometry is particularly bad, since the "lightning-rod effect" results in concentration of the fields. The result is an inverse-square-root divergence of the field (Jackson, 1975, p. 75ff; Monroe, 1980) near the edge of the electrode (see Fig. C.1). In theory this divergence should continue until a distance comparable to the thickness of the electrode, at which it no longer looks two-dimensional. For example, for 0.02mm-thick graphite electrodes 2mm apart, the field should be ten times larger close to the electrode than in the center.

In practice, the divergence is limited by the finite conductance of the sample. As can be seen from Eq. (B.2), any inhomogeneous fields will cause a build-up of space charge will tend to homogenize the field. Actual tests, performing transient measurements on coplanar samples and scanning a focused pulsed-laser spot across the gap, showed little evidence of inhomogeneity (less than a factor of two).

In spite of the apparent insignificance of these effects, the experiments described in this thesis were performed with a "herosandwich" geometry (see Fig. C.2), which combines the advantages of a
coplanar geometry (convenient measurements on bulk samples, large active volume, good optical transmission for TFA measurements) with the well-defined field configuration associated with parallel electrodes. Although the electrodes are not much wider than their separation, the high dielectric constant tends to confine the fields to within the sample; after space-charge build-up, the fields are even more uniform (see...
Figure C.2. "Hero-sandwich" geometry has parallel electrodes as in the normal sandwich geometry, but with a spacing comparable to the width.

Fig. C.3).

Photocurrent-Charging

In the low-temperature TPC measurements there is a serious possibility of changes in the magnitude of the current due to inhomogeneous illumination. The situation is as illustrated in Fig. B.4: since the illuminated region has a higher conductivity $\sigma_L$ than the unilluminated region $\sigma_d$, it tends to short out the voltage drop, which falls across
Figure C.3. Field geometry in hero-sandwich geometry. (a) Before dielectric relaxation occurs, the fields are nonuniform, but tend to be confined to the sample. (b) Following relaxation, the accumulation of space charge on the surface gives a very uniform field distribution.

the unilluminated region. It is easy to show that the field in the illuminated region is reduced by a factor

\[ f + (1-f) \frac{\omega}{\omega_d} \]  \hspace{1cm} (C.1)

where the "filling factor" \( f \) is the ratio of the width of the illuminated region to the total sample width. Unless \( f \) is very close to one, the voltage will fall primarily across the unilluminated region, which will limit the current. More importantly, the change in field will reduce the magnitude of any transient photocurrent excited in the illuminated region.
These experiments cause substantial uncertainty in determining the interaction between a bias beam and the TPC. Such interaction could arise from very interesting microscopic effects, for example re-excitation of trapped carriers by the pulsed light. These measurements therefore provide an important complement to the TPA measurements of the effect of bias. I did perform some of these measurements; unfortunately the geometry was not sufficiently well defined to yield unambiguous results, because of the possibility of charging effects. If these experiments are repeated (they are potentially very interesting), the experimenter should be careful that the results are consistent with the predictions of a microscopic rather than a space-charge model.

C.3 Induced Absorption - Transverse Geometry

For TPA measurements one must align at least two beams relative to one another: the pulsed pump and the probe; this task can be especially difficult if the beams are infrared. In addition, if one is to interpret the magnitude of the resultant signals, at least one of the beams must have a uniform intensity distribution.

This is illustrated in Fig. C.4. If the signal is linear in pump intensity, then the fractional change in transmitted intensity $\Delta I/I$ at a position $\overline{r}$ is just $-\Delta \sigma \phi(\overline{r})$, where $\phi(\overline{r})$ is the number of absorbed photons per unit area and $\Delta \sigma$ is the cross section each photon produces. The total change in light intensity is thus

$$\frac{\Delta T}{T} = \frac{\int \Delta I \delta r}{\int I \delta r}$$

$$= -\Delta \sigma \frac{\int \phi(\overline{r}) I(\overline{r}) \delta r}{\int I(\overline{r}) \delta r}$$

(C.2)
Figure C.3. Well-defined TPA results are possible if either the pump or the probe is uniform. In either case, the nonuniform beam should pass complete within the uniformly illuminated area of the other beam.

If the sample is thick enough that interference effects are not important then the fraction of the incident excitation photons per unit area absorbed in the sample is

\[ a = \frac{(1 + R e^{-\alpha L})(1 - e^{-\alpha L})(1-R)}{1 - R^2 e^{-2\alpha L}}. \]  

(C.3)

Here \( R \) is the fraction of power reflected at each surface \( [(1-n)^2/(1+n)^2] \), \( \alpha \) is the absorption coefficient for the exciting light,
and \( L \) is the sample thickness. For \( aL \gg 1 \) this is equal to \((1-R)\), while for \( aL \ll 1 \) it is \( aL \). For reference, the transmission through the sample is a fraction

\[
\frac{(1-R)^2 e^{-aL}}{1 - R^2 e^{-2aL}}
\]  

(C.4)

of the total.

In order to extract \( \Delta \sigma \), we must know something about the distribution of intensities \( I(\vec{r}) \) and \( \phi(\vec{r}) \). The simplest situation occurs if the excitation beam is uniform. This can be achieved by spreading out the laser light over a large area and passing the light through a small aperture over which the intensity is uniform. The total energy in the pulse \( E^* \) (typically a fraction of a millijoule) must then be measured after the aperture. The areal density of absorbed photons is just

\[
\phi(\vec{r}) = \frac{E^*}{\hbar \omega_x A}
\]

(C.5)

If the entire probe beam passes within the uniformly-excited area, then

\[
\Delta \sigma = - \frac{\Delta T}{T} \frac{A\hbar \omega_x}{aE^*}
\]  

(C.6)

This is the situation used for the experiments in this thesis. Eq. (C.6) defines the quantity \( \Delta \sigma \) even when the signal is nonlinear; in that case the cross section must be interpreted more carefully, however.

An alternative well-defined situation occurs if the probe beam is uniform; then the pump need not be. The pump must then pass totally within the probe spot, and all excited centers are available for probe absorption. In this case \( A \) should be interpreted as the area of the probe beam. This situation is less convenient than the uniform pump, but may be necessary under some circumstances. If neither beam is uniform,
then the signal depends on the particular way in which the intensity
distributions overlap, and the results cannot be related to $\Delta \phi$

**C.4 Induced Absorption - Longitudinal Geometry**

The situation just discussed involved fairly trivial geometrical factors. In contrast, the intensity distribution in the direction normal
to the sample surface can be quite complicated. For example, the formulas (C.3) and (C.4) for the absorption assumed a uniform absorption coefficient throughout the sample thickness $L$. Under some circumstances this is no longer valid. For example, when low-energy excitation light is absorbed primarily by a bias-induced extra absorption, the extra absorption will occur in a region only as thick as the penetration depth of the bias light (see Fig. C.5). In this case formulas (C.3) and (C.4) can still be used, but the factor $e^{-\alpha L}$ must be replaced by

$$\exp\left[-\int_0^L \alpha(x) dx\right]$$

Since the fraction of light absorbed by light traveling in the sample is the same in either direction. Note that the fraction of external light absorbed, $a$, can still be determined from a direct measurement of the transmission.

The inhomogeneity of bias illumination can cause further problems, as shown in Fig. C.5. If the bias photon energy $\hbar \omega_B$ is lower than the energy $\hbar \omega_x$ of the excitation, then all the excitation photons will be absorbed in a region which is uniformly biased; the bias then has a simple interpretation. In contrast, if the bias light is more strongly absorbed, then some of the excitation photons will be absorbed in a region of no bias; one then effectively has two samples in series, and any interpretation must bear this in mind. In principle this problem
Figure C.5. Different Bias Geometries. (a) If $\tilde{\omega}_B < \tilde{\omega}_X$, the sample can be considered uniform. (b) If $\tilde{\omega}_X < \tilde{\omega}_B$, some of the excitation photons may be absorbed in an unbiased region.

could be avoided by always using a sample which is thinner than the penetration depth of the bias light, but this is not always easy to do.
These complications should be kept in mind when bias light is used.

Fortunately, the probe beam is always weakly absorbed, so its dependence on the spatial distribution will always be simple: using Eq. (C.7) in the limit where the integral is small,

\[- \frac{\Delta T}{T} = \int_0^L \Delta \alpha(x) dx \quad \text{(C.8)}\]

where \( \Delta \alpha(x) \) is \( \Delta \alpha \) per excitation times the volume density of excitations at depth \( x \). It is the insensitivity of this quantity to the longitudinal distribution of excitations which legitimates our use of the areal distribution of excitations in the previous section.

Even if there is only one beam, the longitudinal distribution is complicated if there is a nonlinear intensity dependence. For example, the transition to bimolecular recombination is significantly broadened by the exponential penetration of the light: at each depth the onset of recombination occurs at a different time (Orenstein, 1981). In general, it is not possible to establish the dynamics in this situation; one can only check for consistency. To truly establish the intensity dependence one must use thin samples.
APPENDIX D

CABLES

This short appendix provides some useful insight into the behavior of cables by analyzing the cables as transmission lines. Only the transverse electromagnetic (TEM) mode, which has no cutoff, will propagate; because normal signal frequencies are much lower than the cutoff for higher order modes (the longest cutoff wavelength is roughly the circumference of the outer conductor).

For the TEM mode the cable can be represented as in Fig. D.1. Each segment $dx$ of the cable is represented by a capacitance $C^* dx$ between the inner and outer conductors, and a series inductance $L^* dx$ (losses are ignored). The asterisk denotes that the quantities are measured per unit length.

It is easy to show that

$$\frac{dv}{dx} = L^* \frac{di}{dt} \tag{D.1}$$

and

$$\frac{di}{dx} = -C^* \frac{dv}{dt}. \tag{D.2}$$

Combining these, we obtain

$$\frac{d^2v}{dx^2} + L^* C^* \frac{d^2v}{dt^2} = 0. \tag{D.3}$$

That is, the voltage (and also the current) satisfy a one-dimensional wave equation with a velocity

$$v_p = v_g = (L^* C^*)^{\frac{1}{2}}. \tag{D.4}$$

It is less obvious (but nonetheless true), that this velocity is just
Figure D.1 (a) Physical picture for a coaxial cable in the TEM mode. The electric field is directed radially and originates with charge density on the conductors, while the motion of the charge along the cable gives rise to an azimuthal magnetic field. (b) Equivalent circuit for the TEM mode. The electric and magnetic energy storage are represented by a capacitance and inductance per unit length.
the speed of light in the medium filling the cable, \( c/\sqrt{\varepsilon} \).

Eq. (D.3) admits two solutions, a forward- and a reverse- traveling wave. The ratio between the voltage and current for the forward-traveling wave is just the characteristic impedance,

\[
Z_o = \left[ \frac{L}{C} \right]^{\frac{1}{2}},
\]

which can be shown from Eq. (D.1) by inserting the velocity from Eq. (D.4) for \( \partial x/\partial t \). For the reverse-traveling wave the ratio has the same magnitude but opposite sign. The characteristic impedance differs by a geometrical factor (depending only on the ratio of diameters of the conductors) from the impedance in free space, \((\mu_0/\varepsilon_0)^{\frac{1}{2}} \approx 377\Omega\); often \( Z_o \) is 50\( \Omega \).

For very high frequencies the true wave nature of the signals in the cables must be taken into account; it then makes sense to speak of reflections from impedance discontinuities and the like, and to use microwave analysis techniques such as the Smith chart. More commonly, however, the cables are much shorter than the wavelength of the relevant signals in the dielectric. In this limit, the effect of the cables need be taken into account only to first order. It is useful to examine three limiting cases, depending on the load impedance \( Z_L \) with which the cable is terminated.

1. \( Z_L = Z_o \). If the cable is terminated in its characteristic impedance, there are no reflections, and the cable is "invisible" to the source. This is true even if the cable is not much shorter than a wavelength. Note, however, that the cable may not be invisible to the load, if the impedance presented by the source is not \( Z_o \). It is a peculiar but important property of cables that the appropriate equivalent
circuit depends on which end one measures the voltages and currents at.

(2) \(|Z_L| \gg Z_0\). If the terminating impedance is very high, then the voltage will be very large compared to the current entering the load. This requires that there be equal magnitude forward and reverse waves, with their voltages in phase and their currents almost canceling at the load. The phase relationship changes down the cable, such that the current at the source will be larger by an amount determined from Eq. (D.2):

\[ \Delta i = C^*l \frac{dv}{dt} \]  

(D.6)

where \(l\) is the length of the cable. The equivalent circuit of the cable, seen by the source, is therefore a capacitance \(C^*l\) in parallel with the load \(Z_L\), as shown in Fig. D.2.

The high impedance termination is frequently encountered in the laboratory, since many amplifiers have high input impedance. Since the capacitance per unit length is typically 30pF/ft, the effect of the cable can be quite important.

(3) \(|Z_L| \ll Z_0\). For very low impedance, the role of current and voltage is reversed. From Eq. (D.1) we find that the voltage drop along the cable is

\[ \Delta v = -L^*x \frac{di}{dt} \]  

(D.7)

The equivalent circuit for the cable is then an inductance in series with the load.

For \(Z_L\) close to, but not equal to, \(Z_0\), the wave equations (D.3) can be solved with the appropriate boundary conditions. If the impedances are almost matched, the reflected amplitude will be proportional to \(\Delta Z/Z_0\), where \(\Delta Z = Z_L - Z_0\), but the exact expressions are somewhat compli-
Figure D.2. Limiting equivalent circuits for the cable:
(a) $Z_L = Z_0$: invisible to the source. (b) $|Z_L| \gg Z_0$: parallel capacitance $C^* \ell$. (c) $|Z_L| \ll Z_0$: series inductance $L^* \ell$. 
This appendix describes methods of improving signal-to-noise ratio in current measurements. In particular, the "PC box," I constructed for the present research, and which permitted the measurements described in Chapter 6, will be described in detail.

E.1 Basic Principles of Current Measurement

We are concerned with the problem of measuring a small, time-dependent current signal with minimum noise. This problem arises in the PL and PA measurements, because a photodiode acts as a current source, as long as the voltage across the diode is small enough that it remains in its linear region. It is therefore important that the measurement circuit present a small impedance to the detector. The necessary conditions are described, for example, in the very useful photodiode application notes produced by EG&G.

In PC measurements, the sample acts as a current source. This may not be obvious, since the light exposure acts to modulate the conductance \( G \) of the sample. As long as almost all the applied voltage \( V_o \) falls across the sample, however, the current flowing through the sample will be simply \( GV_o \), and the sample will act as a current source. This condition requires that the sample conductance be much less than that of the measuring circuitry. Clearly, for \( V_o \) of order 1000V, failure to meet this criterion would destroy the measurement equipment long before the data became distorted.

The equivalent circuit for both photocurrent samples and photo-
diodes can thus be represented as in Fig. E.1. At normal signal frequencies, the capacitance \( C_s \) dominates the parallel combination, so that 
\[
Z_s = (i\omega C_s)^{-1}
\]
Since \( Z_s \) changes with frequency, it is imperative that the precise value of \( Z_s \) not affect the voltage measured; Thus the impedance \( Z_{in} \) presented by the circuit must always be much smaller than \( Z_s \) at the relevant signal frequencies.

![Sample Diagram](image)

**Figure E.1.** Equivalent circuit for signal sources. The finite resistance is represented by \( R_s \); the capacitance \( C_s \) includes both the sample or junction capacitance and the capacitance of any cables (see Appendix D). The sample is shown connected to a load resistor \( R_L \), which with \( Z_{in} = R_L \) is the simplest current-to-voltage converter.

The simplest method for measuring the current is simply to measure the voltage drop across a known load resistor \( R_L \) as was illustrated in Fig. E.1. For this circuit, the input impedance is simply \( Z_{in} = R_L \), and the measurement is thus limited to times longer than

\[
\tau = R_L C_s \quad \text{(E.1)}
\]

by the criterion that the circuit not load the source. \( \tau \) is simply the
time necessary for the source current \( i \) to charge the capacitor \( C_s \) to the desired voltage \( iR_L \).

![Circuit Diagram]

Figure E.2. Active Load Resistor. The input impedance is smaller than the transimpedance \( V_{\text{out}}/i_p = R_F \) by a factor of the open-loop gain \( A \) of the op-amp.

The circuit in Fig. E.2 overcomes this problem, by using an op-amp to keep the voltage across the resistor \( R_F \) from falling across the capacitor. This configuration is referred to as an "Active Load Resistor," or transimpedance amplifier, or current-to-voltage converter (although the last description also fits a passive load resistor). The voltage out is \(-iR_F\), since the inverting input of the op-amp is held near ground by the negative feedback, and there is almost no current into the op-amp. Thus the ratio of \( V_{\text{out}} \) to \( i_p \), the transimpedance, is \( R_F \), but the input impedance \( v_\text{in}/i_p \) is smaller by the open-loop gain \( A \) of the op-amp. Since \( A \) can be quite large, the feedback resistor can be made substantially larger than \( C_s/i \) before the circuit loads the source.
E.2 Frequency Response of Active Load Resistor

In practice, the operation of the Active Load Resistor is limited by the need to avoid oscillation of the op-amp. An understanding of the causes of oscillation is critical to the proper construction of high frequency circuits. We give here only a simple discussion of the matter; a more detailed treatment is given in Higashi (1982), as well as many textbooks.

The voltage $v_-$ at the input of the op-amp resulting from the feedback voltage $v_{\text{out}}$ at the output is reduced by the voltage divider $Z_s/(Z_s+Z_F)$, where $Z_s$ and $Z_F$ are the impedances of the source and the feedback network, respectively. Since the whole point of the circuit is to allow $Z_F$ to be much greater than $Z_s$, we can assume that this condition is satisfied. $v_-$ appears at the output multiplied by a factor $-A(\omega)$. The "loop gain" $G$ of circuit is the total change around the loop, except for the minus sign: $G = A(\omega)(Z_s/Z_L)$. The steady-state response of the circuit is determined by the condition that the returned voltage $-Gv_{\text{out}}$ be equal to the original $v_{\text{out}}$. This is satisfied only by $v_{\text{out}} = 0$, unless $G = -1$, in which case a signal of arbitrary amplitude is allowed. In other words, oscillation can (and will) occur if there is any frequency at which $G(\omega) = -1$.

Normally the op-amp is operated in a region in which the open-loop gain is falling inversely to the frequency, and has a 90° phase shift: $A(\omega) = (\omega_C / \omega)$. As mentioned before, the source capacitance usually determines $Z_s = (i\omega C_s)^{-1}$. Thus oscillation will occur if at some frequency $\omega Z_L$ is equal to

$$R_{F,\text{max}} = \frac{\omega_C}{\omega^2 C_s}$$  \hspace{1cm} (E.2)
\[
\frac{(\omega GB)^2}{C_s}
\]

(E.3)

Note that the equality must include phase as well as magnitude.

Thus, the maximum transimpedance is proportional to the square of the desired time scale \( \tau \), in contrast to the simple load resistor, for which it is proportional to \( \tau \) (see Eq. (E.1)). The source capacitance is still a limiting factor; any reduction of this capacitance, for example by using shorter cables, will be proportionately rewarded. In addition, \( R_{F,\text{max}} \) is directly proportional to the gain-bandwidth product \( \omega GB \) of the op-amp, so it is critical that the fastest possible op-amp should be used. The choice of op-amp also depends on noise considerations, however; as discussed by detail by Higashi (1982, Appendix I), and on more subtle oscillation problems discussed below.

In practice, to avoid ringing \( R_F \) must be chosen to be somewhat smaller (say, a factor of two) than the \( R_{F,\text{max}} \) corresponding to the desired time response \( \tau \). With this feedback resistor, oscillations would still occur, but at a slightly higher frequency than \( 1/\tau \). This is avoided by putting a capacitor \( C_F \) in parallel with \( R_F \); this changes the phase of the loop gain by 90°. The capacitor is chosen so that \( R_F C_F = \tau \). The circuit response is then predominantly a one-pole response determined by the feedback network; the next pole is not far away, however.

**E.3 Physical Realizations of the PC Box**

In this section I will describe the transimpedance amplifiers I have built and which are still in the Kastner lab. I will not discuss the amplifiers associated with photodetectors, but rather the "PC boxes" used for making photocurrent measurements. There are two incarnations of the PC box, the first in a in a cheap aluminum box, the second in a wel
shielded box with a large number of switches on the outside. Other differences are discussed below.

The basic circuit of the PC Boxes is illustrated in Fig. E.3. Functionally, the circuit is simply an active load resistor. For flexibility, however, there are many possible feedback resistors $R_F$ and capacitors $C_F$. These are chosen to achieve the maximum transimpedance without oscillation at the desired time constant; the optimum settings depend on the source impedance.

Figure E.3. Circuit diagram for the PC boxes. In addition to the functional element $R_F$, $C_F$ is added for stability (as is $C_B$ in the first incarnation), and the op-amp is protected by a series resistor.

Note that, since the capacitance and resistance are chosen independently, it is possible to choose values such that the circuit oscillates. This is the principal difference between the PC box and the
detector amplifiers; for the latter, the source impedance is known and unchanging, and the capacitors and resistors are generally switched together. The PC Box thus requires a greater awareness of oscillation issues by the operator as well as the designer of the device.

When using the PC box in a new application, the optimal switch settings for every time range should be determined in advance by looking at some well-known signal. The appropriate procedure is to start with an unnecessarily large $C_F$ for each load resistor, gradually remove capacitance until oscillation or incipient oscillation is observed, then add a little capacitance. Note that incipient oscillation will be manifested as "ringing" in the time response to a pulse, so the box should not be operated even close to oscillation.

Furthermore, sustained oscillation may irreversibly damage the op-amp by overheating it. The first incarnation of the PC box, which used an Analog Devices Model 50J op-amp suffered many fatal incidents as a result of oscillation. Since the 50J costs about $200, this was unsatisfactory, and the second incarnation uses a PMI OP-37, which costs about $10. These devices were chosen primarily on the basis of their gain-bandwidth product. The noise analysis was incomplete, however, and any further improvements should carefully consider noise sources, including current noise. Gregg Higashi recommended the Analog Devices Model 45 for its low current noise. The circuit board used in the second incarnation is compatible with the Model 45 op-amp as well as the OP-37.

At long times, one can use quite large feedback resistors—of order 10MΩ. Very small capacitances are then sufficient to prevent oscillation; in fact, stray capacitances tend to degrade the time response of the circuit for the large load resistors. For this reason, the second
incarnation uses cold-switched, ultra-low capacitance FET switches (Signetics SD 5000 series). These switches are designed to change between on and off very fast, which is totally unnecessary in this application. It may well be that much cheaper switches would have the necessary low capacitance between source and drain. The switches are presently driven by 15V signals; for a slight sacrifice in on resistance it would be easy to drive the PC box using 5V external digital signals. This would also permit external gating of the device.

Both boxes were constructed using ground planes close to the circuit elements. This is good RF practice, since it minimizes both stray inductance and stray capacitance to other leads besides ground. Unfortunately the large number of devices switched prevents the compact construction desirable in an RF circuit. The two-sided PC board used for the second incarnation, which was produced in the LNS electronics area on the fourth floor of MIT's Building 26, is shown in Fig. E.4.

Both boxes include RF filtering of the power-supply leads. Both inductive and capacitive filtering are used, to filter out interference from both low and high impedance sources. Examination of op-amp spec sheets will show that high-frequency signals on the power supply can easily find their way into the output of the circuit. For the second incarnation, the leads are filtered using RFI filters purchased with the shielded box from Compac corporation.

In a further effort to reduce stray capacitance, in the second incarnation the switches for the capacitors and for the smaller value resistors are not even connected to the input unless another switch is closed. This reduces the capacitance seen by the circuit in the time ranges when strays are important, but the arrangement used may not be
Figure E.4 Circuit board (two sides) used for second incarnation of PC box.

optimal.

The boxes include some features that do not appear to be necessary to its operation. One of these is a resistor in series with the input. This should serve as a current limiter to protect the input from excessive voltages. It will cause some degradation of the signal if it becomes larger than the input impedance of the rest of the circuit; this condition should be avoided. A feature used in the first incarnation is a capacitor connecting the output pad to ground. The need for this was established empirically: it was required if oscillation was to be avoided. This only makes sense is if the output impedance of the 50Ω is
not sufficiently low and has a phase shift that contributes to the overall loop gain.

Both PC Boxes work well at long times, the second incarnation being less prone to damage, better shielded, and more reliable. At high frequencies, however, the second box fails to achieve what it would seem to be capable of: to avoid oscillations, large capacitors must be used, degrading the frequency response. The reason for this is somewhat subtle, and has to do with the frequency response of the OP-37 op-amp, illustrated in Fig. E.5. The magnitude of the open-loop gain drops steeply somewhat above $|A|=1$. Such a variation is always associated with a phase that is closer to 180°. In the language of electrical engineering, the "phase margin" (difference of the phase for 180° at $|A|=1$) of this amplifier is poor. As a result, substantially more capacitance must be put into the feedback loop to counteract this phase shift. The designers have have made a moderately good op-amp perform better at low frequencies by including insufficient compensation for high-frequency operation; they have also redefined "phase margin" to mean the margin at $|A|=1$, the minimum recommended value. This example shows the importance of careful examination of amplifier characteristics in evaluating their potential performance.

E.4 Further Improvements

The devices described here have been very useful. Particularly at long times they provide for substantially better transimpedance than a simple resistor, and this made possible the measurements discussed in Chapter 6. Nonetheless, the large capacitance of the cables going to the bottom of the cryostat (~400pF) limits the measurement much more than seems necessary.
Figure E.5. Open-loop Gain of PMI OP-37 op-amp. Note the steep slope of the characteristic near $\omega_{GB}$, hinting at the poor phase margin shown on the right.

One way to overcome this would be to introduce a new cryostat, with substantially shorter cables. This of course sacrifices the flexibility of the present cryostat, but should be considered. Another option is to provide for an amplifier that functions in the cryostat. This could be done by mounting an amplifier in the vacuum space. Alternatively, an amplifier might be constructed that operates even at low temperatures. Presumably this would be a JFET-based device, similar to those used (for low noise) on space probes. Lyman Hurd in Ranier Weiss's astrophysics group at MIT has some experience with such devices. It is of course imperative that the frequency response be flat for all temperatures. The most difficult task would be to allow switching between different circuit configurations, which is probably necessary to optimize the measurement for different time scales. But it might be done.
Figure E.6 Active Shielding. By driving the cable shield with a follower, the sample is relieved of the burden of charging the cable capacitance.

The most realistic possible improvement would use an Active Shield, or "guard," in addition to an active load resistor. The principle is illustrated in Fig. E.6: The limitation on the present system arises because the sample must charge up the cable capacitance, the outside of which is at ground potential. By using double-shielded cable, one can actively keep the inner shield at the same potential as the inner conductor; the follower then must charge up the capacitance between the two shields. By this technique, the capacitance of the cable can be made effectively invisible to the sample, to the extent that the follower
accurately follows the signal. I tried this technique on the bench, and it works rather well in conjunction with a simple load resistor; for unknown reasons it did not improve the time response when used in conjunction with the PC Box. It is unlikely that this is a fundamental problem, but it is important to understand the equivalent circuit of the system very well. The Active Shield approach seems to be the most promising for further improvements in the PC measurement.

A final possibility is to use an op-amp at frequencies greater than its bandwidth. This results in an integration of the signal, which may be useful; it may also be possible to differentiate the result to reproduce the signal, if the amplifier response is well-defined. This latter technique produces no improvement in signal to noise, except if the current noise from the feedback resistor itself is limiting the measurement. Under these circumstances, the noise is reduced as the inverse square root of the feedback resistor.
The rapid increase in the use of transient experiments over the past few years is partially the result of the availability of cheap, convenient, digital, data acquisition equipment (the reverse is also true). The discrete nature of the digitization process, illustrated in Fig. F.1, necessitates special techniques in order to optimize performance, which differ from standard analog techniques. A few of these tricks are included in this appendix.

F.1 Time Resolution

One problem associated with digital acquisition is the finite time interval $\delta t$ between data samples. The most striking manifestation of this limitation is "aliasing": a high frequency sinusoidal input (with a period shorter than twice the sample interval) will result in an output which is also a sinusoid, but the new sinusoid will have a very different frequency, which is the difference between the true frequency and the closest harmonic of the sampling rate. Although this constitutes very severe distortion of the output, it is usually not important for acquisition of transients, which do not have a repetitive waveform; nonetheless, it is obviously impossible to obtain any information on a time scale faster than $\delta t$.

The other problem associated with the finite time resolution is associated with signal-to-noise ratio (S/N). It results from the fact that the sample interval $\delta t$ corresponds only to the spacing between samples; the actual time over which the data is acquired (the sample dura-
Figure F.1. Digitization of Signal. An input waveform, has a continuously varying voltage as a function of time. The digitizer samples this voltage at regular intervals and reports the nearest multiple of $\delta v$ to the true voltage, delivering a waveform which is discrete in time and voltage.

tion) is generally much less, and comparable to the smallest available sample interval (2ns for the Biomation 6500). Since the effective noise bandwidth is governed by the sample duration, while the signal bandwidth is governed by the sample interval, severe $S/N$ degradation occurs. In
time domain language, the instrument samples only a small piece of the waveform plus noise during the interval $\delta t$; by averaging over the whole interval, the relative noise can be substantially reduced. This averaging can be achieved only by low-pass filtering the signal before it reaches the digitization stage. The time constant of the filter should be comparable to the sample interval in order to retain maximum information in the recorded waveform.

It should be noted that filtering of power-law signals presents some special problems, because of the large spike at $t=0$. For power laws, $i(t) \propto t^{-\eta}$, the time necessary before the response is accurate to 1% is roughly $\frac{5\tau}{1-\eta}$. Thus the normal $5\tau$ rule is changed to a $10\tau$ rule for a $t^{-\frac{1}{2}}$ dependence. For a $t^{-1}$ decay the area under the spike is divergent, so in theory one must wait forever. The response is also sensitive to higher-order poles in the frequency response. Some of these problems are discussed by Orenstein (1981).

F.2 Analog-to-digital Conversion

The heart of any digital data acquisition system is an analog-to-digital converter, or ADC, which produces a set of bits as a digital representation of the analog input voltage. Although other designs are sometimes found, most ADCs are simply digital-to-analog converters (DACs) together with a comparator. The ADC control circuitry compares the output of the DAC with the signal to determine the bits, and changes successively less significant bits to get the closest match.

There is an important tradeoff between the speed of an ADC and its resolution (specified in bits). For example, the Biomation 6500 has a 2ns sample time but only six bits of resolution, because it uses a "flash converter" which has sixty-four comparators in it. The input is
represented by a six bit binary number, so the voltage resolution $\delta V$ is only $1/64$th of full scale. To take full advantage of this limited voltage resolution, of course, the signal should be amplified so that it is as large as it can be while avoiding overloading or offscale readings.

ADCs with high resolution suffer from a different problem. The output of an ADC can be no more accurate than $\delta V$, but it is often significantly less accurate. For example, if the most significant bit of the DAC generates a voltage differs by a part in $2^N$, then only $n$ bits of the ADC output will be worthwhile. As a rule, ADCs with more than ten bits resolution should be carefully examined to be sure that they are as accurate as they are precise. Fig. F.2 illustrates the problem.

**F.3 Averaging**

After digitization, the signals are numerically averaged (in the case of the Nicolet, the algorithm is only approximate) over many repetitions of the experiment (laser pulses). The signal is the same on each pulse but the noise has random positive and negative sign; The S/N ratio there increases as $N^{1/2}$, where $N$ is the number of pulses.

This calculation assumes that the noise on successive pulses is uncorrelated, in other words, that the correlation time of the noise is shorter than the time between pulses. Since the pretrigger capability allows subtraction of a time-varying baseline, only variations on which occur between the baseline and the signal are important. The signal is thus effectively high-pass filtered with a time constant comparable to the time scale of the measurement. The important frequency components of true noise will be uncorrelated on the time scale of the repetition time, and will be effectively averaged.
Figure F.1. Accuracy of ADC. The response of one channel of the Nicolet 4094 to an accurate ramp voltage. The flat response near the center is a result of errors in conversion. The maximum error is approximately 4.6 μV (2 bits).

Various sources of pickup are not so well averaged, however. Sixty-cycle pickup will be important for long time measurements, because the frequency is comparable to the repetition rate. Also, high-frequency pickup with a very stable frequency (for example, a radio station) will average poorly.

Most importantly, interfering signals which are synchronized with the signal (for example radio-frequency interference, or RFI, from the laser) will not get smaller at all as a result of averaging. Synchronous noise can be removed only by taking data with the actual signal removed and subtracting, with a $\sqrt{2}$ loss in S/N. Because of the poor averaging
characteristics of pickup, it is generally worth the time and effort to reduce them to their minimum size before taking data.

**F.4 Dithering.**

Averaging increases the accuracy with which data is acquired; the minimum resolvable voltage shrinks as $N^{-\frac{1}{2}}$. Eventually this noise-limited resolution is equal to the digitizing resolution of the ADC. At this point one might think that the ultimate in performance had been achieved for the equipment used. In fact, under appropriate circumstances, one can achieve substantially finer resolution than this.

In order to determine the prerequisites for achieving such accuracy, we need only consider the averaging process for a single time bin. On the $k$th transient the input signal $v_k$ is the sum of the true signal $v_o$ (which is the same for all $k$) and the noise, which varies from pulse to pulse:

$$v_k = v_o + \Delta v_k$$  \hspace{1cm} (F.1)

Pulse-to-pulse fluctuations in the magnitude can be included within the noise term, with a magnitude proportional to the signal at the time in question.

The digitization process produces an integer $I_k$:

$$I_k = \text{RND} \left[ \frac{v_k}{\delta v} \right]$$  \hspace{1cm} (F.2)

where the rounding function RND produces the value rounded to the closest integer. For example, the true desired input $v_o$ is represented by

$$I_o = \text{RND} \left[ \frac{v_o}{\delta v} \right]$$  \hspace{1cm} (F.3)
The result of averaging \( I_k \) is just

\[
\bar{I} = I_o + \frac{1}{N} \sum_{i=1}^{N} \text{RND} \left[ \Delta I + \frac{\Delta v_k}{\delta v} \right]
\]  

\[ (F.4) \]

where

\[
\Delta I = \frac{v_o}{\delta v} - I_o
\]  

\[ (F.5) \]

is the extent to which \( I_o \) fails to represent \( v_o \).

If \( \bar{I} \) is to accurately represent the true value \( v_o \), then the sum in Eq. F.4 must approach \( \Delta I \) as \( N \) becomes large. Clearly, a noise-free signal, having \( \Delta v = 0 \) will not satisfy this criterion, since the rounding process will always give zero. Only if the noise \( \Delta v \) is comparable to \( \delta v \) can accurate results be obtained. It is therefore possible to be in the paradoxical situation of having to add noise to improve the accuracy of the averaged signal. This process of adding noise is called dithering, and its effectiveness depends on the character of the added noise, as we now demonstrate.

Since we are looking only at a single time bin, the only characteristic of the noise (added or otherwise) of concern is its probability distribution function, or PDF. The PDF \( P(\Delta v) \) describes the probability that the noise will have a given amplitude \( \Delta v \), so the various \( \Delta v_k \)'s are chosen at random from this distribution. Two important PDFs are that for white noise, which is a Gaussian, and that for a sinusoid, which has the form \((A^2 - \Delta v^2)^{-\frac{1}{2}}\), where \( A \) is the amplitude. These forms are illustrated in Fig. F.3. The sharp peaks (inverse square root divergences) at the extrema of the sinusoidal PDF result from the zero derivative of the signal near its maximum, and these play an important role in what follows.
Figure F.3. Representative probability distribution functions. (a) The Gaussian PDF appropriate for white noise. (b) The PDF for a sinusoid. Note the sharp peaks near the extrema.

The ultimate limit on the resolution is determined by averaging over an infinite number of pulses, at which point all possible values of the noise have been observed in proportion to the PDF. The error is then

\[
\text{Error} = \delta v \left[ \int_{-\infty}^{\infty} P(v) \left( \text{RND} \left\{ \Delta I + \frac{\Delta V}{\delta V} \right\} - \Delta I \right) dv \right]
\]  

\[(F.6)\]
This error depends on the initial error $\Delta I$. For example, the error is zero for $\Delta I = 0$ or $\frac{1}{2}$, independent of the magnitude of the noise, since for a symmetrical PDF the values of $I$ delivered are symmetrically arranged around the true value. In Fig. F.4 are plotted the error for $\Delta I = 0.25$, which is representative of the general level of error, as a function of the magnitude of the noise. Note that for Gaussian noise, the error rapidly approaches zero once the amplitude becomes comparable to the resolution $\delta v$. For the sinusoidal PDF, in contrast, the effective resolution decays much more slowly and one must use a much greater amplitude to achieve a given accuracy. This slow convergence is a result of the sharp peaks in the PDF, noted earlier. To the extent that most of the voltage is within $\delta v$ of the peak, the dithering is ineffective. Because the total probability of being within $\delta v$ of the maximum goes as $(A/\delta v)^{-\frac{1}{2}}$, the envelope falls off only as $A^{-\frac{1}{2}}$.

Fig. F.5 shows the optimum $S/N$ and the requisite dithering amplitude for a finite number of samples, determined by equating the digital error with the rms statistical noise. Clearly the Gaussian is far superior. It is for this reason that Jim Stathis, in constructing the Digital Dithering Device for use with the Biomation, went to the effort of producing a Gaussian PDF rather than the much simpler-to-obtain sinusoid. The ultimate resolution typically goes as $\delta v/N^{\frac{1}{2}}$. This is substantially better than the apparent resolution limit of $\delta v$ set by the ADC, but substantially poorer than $\delta v/N$ which could be achieved by analog addition of the signal and conversion to digital with the same resolution. The latter process would require as much resolution in the ADC as in the memory, however, and is quite impractical. Dithering, on the other hand, is quite practical and even necessary for achieving reason-
Figure F.4. Representative error for the two PDF's shown in Fig. F.3. The abscissa is the RMS noise, which gives an indication of the statistical error introduced when only a finite number of pulses are averaged. The ordinate shows the ultimate minimum error. The zeroes in the error for the sinusoid will be correct only for the particular ΔI chosen (.25); the envelope (dashed line) is more representative.

able performance with low resolution ADC's.

It is important to be able to recognize the effects of digital error in an averaged waveform, because the signal is distorted differently than with typical analog problems. Fig. F.6 illustrates the effects of digital error on a typical power-law-type transient. The top
Figure F.5. Minimum achievable noise. For a finite number of sweeps, one should choose the amplitude of the noise to minimize the sum of the statistical noise and the intrinsic noise. The minimum noise achieved as a result of this tradeoff is shown here.

The figure shows the input signal, or equivalently the result of averaging with sufficient noise. The bottom figure shows a single digitized waveform, or equivalently the result of averaging with no noise. The center figure is intermediate; clearly the digitization has produced drastic effects, although there is enough noise that the clear steps have been rounded somewhat. This is the most dangerous situation, because one may be tempted to interpret the features; this is especially likely once the data is plotted logarithmically. There is usually little digital error in the large parts of the transient, because the effective noise due to pulse-to-pulse magnitude fluctuations is largest there.
Figure F.6. Effect of digital error on a waveform, described in the text.
Figure F.7. Log-log plot of averaged noise. To produce this plot, simulated Gaussian noise was averaged using a constant fractional time interval $\Delta t/t$. The result clearly resembles a $t^{-\frac{1}{2}}$ power law, although there is really no data.

Once the signal has been correctly averaged and transferred to the computer, it would seem that no further difficulties could arise. One further pitfall lurks, however, which is associated with the averaging of power-law-like data over time, the normal practice of plotting such data on logarithmic axes, and the fact that one often takes data with
Figure F.8. Similar to Fig. F.7, except that data decaying as $t^{-0.8}$ is added to the noise. The S/N ratio is one in the center of the plot; for shorter times, the decay looks steep, for longer times shallow.

marginal S/N ratio.

In a power law, the signal decays by the same fractional amount for equal increments of log time, and each such increment contains the same amount of information about the decay. It is therefore useful to average the data over time intervals $\Delta t$ such that $\Delta t/t$ is some constant fraction (typically 15% for my data). Of course, for an exponential decay this will cause distortion once the time becomes longer than the exponential time constant, but this problem has a distinctive signature in that the averaged signal decays by more than a factor of two between successive averaged time points. For power law decays, however, the variation in
the data over the averaging interval will be negligible as long as the power law is less than around $t/\Delta t$, or about 6 for a 15% fraction.

A more serious problem arises in the averaging of noisy data, a situation which is quite common. Since the number of points per averaging interval is proportional to the time, the RMS noise (assuming white noise) falls as $t^{-\frac{1}{2}}$. Fig. F.7 shows this effect, which is the average of only noise, with no signal. The unwary experimenter might conclude that this was a $t^{-\frac{1}{2}}$ power law, albeit a noisy one. Similarly, the averaged version of $t^{-0.8}$ plus noise is shown in Fig. F.6; the curve appears to change slope, when in fact the true data is $t^{-0.8}$ throughout.

In order to avoid these problems, one must be rather conservative about what data one interprets. Specifically, data with a S/N approaching one should be totally ignored (preferably not even plotted to avoid unintentional bias). This criterion is obvious on a linear plot, but with logarithmic vertical axes the data will typically extend only over a factor of two, and this may not look like much. It is also important to have a means of seeing any negative data. For example, my data-plotting routines (used to produce these figures) place points on the bottom axis to indicate data that are off of the graph.
ACKNOWLEDGEMENTS

For me, the true meaning of life is found through interaction with other people; this is as true in science as in any other part of life. I am grateful to have been part of such a wonderful "family" during my time at M.I.T.

My early days were stimulated by the influence of Joe Orenstein and Gregg Higashi. Joe shared his physical insight and also his dry humor. I am grateful to have been around for the great Multiple-Trapping Revolution, which was successful because of the simplicity and accessibility of the ideas; this has inspired my own work as well. Joe has been a great role model for me in other ways as well, and I am glad to have had some opportunity to know him.

Gregg taught me physics, but more important numerous laboratory techniques, which can be acquired only through personal interaction. His didactic style has helped give me an awareness of the importance of communication, and inspired me to similar attempts at passing on knowledge; in a very real sense his personality is still propagating through our lab. More importantly, Gregg also showed me a sense of morality and of personal responsibility in interaction with people, both in the laboratory and in "real life," that I hope I always retain. He also showed me it was possible to drink a lot of beer and still get work done, for which I will forever curse him.

Larry Robins has been a helpful labmate for several years, as well as a companion on trips as far as Tokyo. He has grown tremendously in this time, becoming much more relaxed with people. His humor has lightened many an uneventful afternoon or evening.

Dwight Gibson has shown an willingness to enquire into the most
fundamental aspects of our understanding that really affirms our identity as physicists, and not just scientists or technicians. Fortunately for him, he has tempered this willingness with a desire to advance knowledge, but I hope the fundamental questioning always remains with him, and with me.

Tineke Thio has the inspired vision of the "Renaissance Man" (sic), in which the world of Art and the world of Science are really the same World. I have yet to see the vision in anything close to its entirety, but even the rare glimpses are wonderful to behold.

Jim Stathis has been my office mate for three years, and is without a doubt the perfect office mate: undemanding, but there when you want to talk about something. I couldn't count the times I've been puzzled by something which got resolved in a conversation with Jim. He has an openness to undefined conversation, an appreciation for the magic of intuition, that is common to all great minds. He is also an excellent critical thinker; I have the greatest of respect for him, and I'm glad we've shared an office. I'm still waiting for him to realize how profound his awareness really is.

Rob Kwasnick was my friend and office mate for many years. Although I sometimes felt our differences were to vast to be bridged, he perpetually made the extra effort. He helped my ability to communicate more than anyone I know, both by the example of his own willingness to make a statement when things were less than certain, and through our many hours of conversations. In addition, he never shrunk from analyzing what was happening, even while it was happening. While this sometimes destroys the happening, it also provides an awareness which cannot be obtained in any other way.
Finally, the most profound debt of all I owe to Marc Kastner. It would be impossible to list all of the things he has taught me about physics and about life as a physicist. This is just a reflection of his willingness to share his great insight; I cannot count the times I have simply mentioned a topic to Marc and been enriched in a few odd minutes with an overview of a subject that would take weeks to wrest from books. This applies equally well to insights about life, except that these cannot be gotten from books at all.

More importantly, I have acquired (or had confirmed) from Marc a conviction that the beauty of the physical world is often comprehensible in a simple way; in fact, its beauty is its simplicity. This is not just a statement of outlook; Marc demonstrates it with his work. This is one lesson I have taken in completely, and I will continually be looking for the simple viewpoint on complex problems.

Marc's most outstanding feature, however, is his respect for his students as people. He is genuinely concerned for our well-being and personal growth above any benefit for himself; and this is reflected in tremendous growth among his students. It's no accident that Marc has many excellent students, for not only does he attract bright students to begin with, not only does he train them well, but he helps them to grow into complete scientists. His influence will last throughout my life, and beyond.
REFERENCES


G. Brunst and G. Weiser, "Defect Controlled Conductivity in \text{As}_2\text{Se}_3 Single Crystals"


Lynda E. Busse, 1984, "Temperature Dependence of the Structures of \text{As}_2\text{Se}_3 and \text{As}_2\text{S}_3 Glasses near the Glass Transition," Phys. Rev. B 29, 3639.


Phys. 11, 35.


J.M. Hvam and M.H. Brodsky, 19??, "Experimental Evidence for Trap Controlled Dispersive Transport in Phosphorus Doped Hydrogenated Amorphous Silicon."


M.A. Kastner, 1984, "Defects in Chalcogenide Glasses"


H. Michiel and G. J. Andriaenssens, "Analysis of Dispersive Transport by Discretization of a Continuous Distribution of Localized States"
Don Monroe, "Hopping in Exponential Band Tails," to be published.
Don Monroe, "Resolution of the Conductivity-Thermopower Conflict in Amorphous Semiconductors," to be published.
Don Monroe and M.A. Kastner, 1984, "Transient Photo-Induced Absorption Spectroscopy in a-As$_2$Se$_3$ in the presence of Strong Bias


L.H. Robins and M.A. Kastner, 1984, "Recombination and Excited-State Absorption at Photoluminescence Centres in Crystalline and
Amorphous Arsenic Triselenide, Philos. Mag. 50, 29.


M. Roilos and E. Mytilineou, 1974, "Hall Mobility in $\text{As}_2\text{Se}_3$ and $\text{As}_2(\text{Se,Te})_3$ Glasses," in Amorphous and Liquid Semiconductors, edited by J. Stuke and W. Brenig (Taylor and Francis, New York, ) p. 319.


F. Urbach, 1953, Phys. Rev. 92, 1324.


G. Weiser, "Chemical Bonds and Electronic States in Monoclinic As$_2$Se$_3$"


CONFERENCES
The following references are the proceedings of conferences:


Proceedings of the Fifth International Conference on Liquid and Amorphous Semiconductors (Garmisch-Partenkirchen, 1973):

Proceedings of the Eighth International Conference on Liquid and Amorphous Semiconductors (Cambridge, MA, 1979):

Proceedings of the Ninth International Conference on Liquid and Amorphous Semiconductors (Grenoble, 1981):

Proceedings of the Tenth International Conference on Liquid and Amorphous Semiconductors (Tokyo, 1983):


Proceedings of the International Topical Conference on Optical Effects in Amorphous Semiconductors (Snowbird, Utah, 1984):
Optical Effects in Amorphous Semiconductors, ed. by P.C. Taylor and S.G. Bishop (American Institute of Physics, New York)