AD-A048 194 NORTH TEXAS STATE UNIV DENTON DEPT OF PHYSICS F/G 20/12
OPTICALLY INDUCED HOT ELECTRON EFFECTS IN SEMICONDUCTORS.(U)
AUG 77 A L SMIRL N00014-76-C-1077
UNCLASSIFIED NL

1 OF 3
AD A048194

(12) F.G.

Annual Summary Report

— Optically Induced Hot Electron Effects in Semiconductors

prepared by
Arthur L. Smirl
Department of Physics
North Texas State University
Denton, TX 76203
September 1, 1976 - August 31, 1977

Prepared for the Office of Naval Research
Electronic and Solid State Sciences Program
Under Contract NR 318-048

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle) (6)
Optically Induced Hot Electron Effects in Semiconductors,

5. TYPE OF REPORT & PERIOD COVERED
Sept. 1, 1976 - Aug. 31, 1977

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s) (10)
Arthur L. Smirl

8. CONTRACT OR GRANT NUMBER(s)
N00014-76-C-1077
NEW

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Physics
North Texas State University
Denton, TX 76203

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (15) NR 014-76-C-1077

11. CONTROLLING OFFICE NAME AND ADDRESS
ONR, Electronic and Solid State Sciences Program
Office of Naval Research
Arlington, VA 22217

12. REPORT DATE
August 21, 1977

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
ONR Resident Representative
528 Federal Building
300 East 8th Street
Austin, TX 78701

(11) 21 Aug 77

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

(12) 296p.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
Approved for public release; distribution unlimited

(9) Summary rept. 1 Sep 76 - 31 Aug 77,

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Photoexcited Carriers
Germanium
Ultrafast Phenomena

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We ~~report here, our~~ progress* in using the high electric fields and the ultrashort pulses that can be derived from Nd:glass and CO_2 lasers to investigate hot electron effects in semiconductors. This research is directed toward obtaining fundamental information concerning the saturable optical transmission properties and hot electron dynamics of semiconductors on a picosecond time scale and the stress and electric field biasing of the absorption properties of semiconductors under intense short pulse CO_2 laser radiation.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

* is reported 4/09 923 mt

FOREWORD

This Summary Report by Arthur L. Smirl, Department of Physics, North Texas State University, Denton, Texas, covers research progress for the period September 1, 1976 to August 31, 1977, under Office of Naval Research Contract NR 318-048. N00014-76-C-1077

The use of trade names in this report does not constitute an official endorsement or approval of the use of such hardware or software. This report may not be cited for the purposes of advertisement, and its publication does not constitute Navy approval of findings or conclusions.

Arthur L. Smirl
Principal Investigator

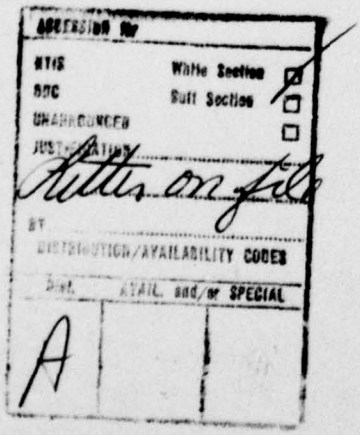


TABLE OF CONTENTS

SUMMARY

In this work, we report on our progress in attempting to use the high electric fields and the ultrashort pulses that can be derived from Nd:glass and CO_2 lasers to investigate hot electron effects in semiconductors. Intense ultrashort optical pulses having durations of a few picoseconds and peak powers of 10^8 watts can be readily generated by mode-locking a Nd:glass laser. The extremely high power and short duration of pulses from these lasers make possible the study of saturable optical transmission properties and hot electron dynamics of semiconductors on a picosecond time scale. In addition to this study of carrier dynamics, we are also attempting to investigate stress and electric field biasing of the absorption properties of semiconductors under intense, short-pulse CO_2 laser radiation.

Hence, we report on the development (with coworkers at the University of Arizona) of an initial model (see Appendix A) that describes the generation and the temporal evolution of hot electron-hole plasmas produced in the semiconductor germanium by the absorption of intense, ultrashort optical (λ = 1.06 μm) pulses. This model is shown to be in agreement with the experimentally-observed enhanced transmission of single, energetic pulses through thin germanium samples as a function of incident pulse energy for two sample temperatures, 105 K and 300 K. It has also successfully described the transmission of a weak probe pulse through the germanium as a function of time delay after an energetic excitation pulse for different excitation energies and for sample temperatures of 105 K and 300 K. Subsequent to the development of this model, van Driel and coworkers conducted

further transmission studies (see Appendix B), in which the energy band gap of the germanium was tuned by hydrostatic pressure, that corroborated this proposed model.

Recently, however, the optical pulsewidth dependence of thin germanium samples was experimentally measured, as suggested in our earlier work (see Appendix C). We found these measurements to be in substantial disagreement with the predictions of our model. The question then arose as to whether this disagreement with the proposed model was due to assumptions made in deriving the model, experimental uncertainties in physical constants used in the calculations, or to relevant physical processes omitted from the model. A numerical investigation (Appendix D) of the effect of the electron-phonon coupling constants on the model revealed that, given the uncertainties in these and other constants, the pulsewidth experiment was not a definitive test for the model. In fact, given the uncertainties in certain of the physical constants contained in the calculations, many of the questions concerning the physics of ultrafast phenomena in germanium can only be answered by further experimental investigation.

ULTRAFAST TRANSIENT OPTICAL RESPONSE OF ELECTRON-HOLE PLASMAS IN SEMICONDUCTORS

Theory

During the past year, the theoretical activity in this project has been a detailed study of the theory for the transmission of ultrashort optical pulses through germanium. At the beginning of the year, an initial model had been developed by A. Elci, A. L. Smirl, J. C. Matter and M. O. Scully. Smirl and Matter had also completed single pulse and excitation-probe pulse experiments with germanium. In addition, an attempt had been made to generate numerical solutions to the complicated coupled set of integro-differential equations in the theory.

However, there were several problems in the original computer codes. The computer program was lengthy and expensive to execute, and the solution to the differential equations was crude. Several modifications were needed to improve the accuracy and execution time.

There were four areas in which the numerical techniques were improved. First, the execution time was considerably lowered by using Gauss-Laguerre quadrature formulas to evaluate the integrals. Next, the accuracy of the solution to the differential equations was improved by using a fourth order Runge-Kutta solution. The third major area of improvement was the intialization routine which starts the program. By including limiting valves calculated in the high temperature limit, the executiom time was improved by nearly an order of magnitude. Finally, a correction routine was included

in the program to check for and correct for roundoff error due to the large number of iterations required to complete the calculation.

Having developed an efficient and accurate computer routine, W. P. Latham and A. L. Smirl generated the final theoretical curves comparing the experiments of Smirl and Matter to the theory of A. Elci, et al. The results were presented in the paper by A. Elci, A. L. Smirl, J. C. Matter and M. O. Scully.[1] A copy of this paper is included as Appendix A of this report.

By the time the paper by Elci, et al.[1] was completed, additional optical experiments, in which the band gap was tuned by pressure, had been performed by H. M. van Driel, J. S. Bessey and R. C. Hanson.[2] The computer code was modified to include these pressure effects and the corresponding theoretical curves were generated by Latham and Smirl. The agreement between these new experiments and the proposed model was excellent. The pressure experiment is reviewed in the preprint contained in Appendix B of this report.

Subsequently, J. S. Bessey, B. Bossachi, H. M. van Driel and A. L. Smirl[3] completed single pulse transmission studies in which the incident optical pulsewidth was varied. (A preprint of this paper is included in Appendix C of this report.) The pulsewidth experiment indicated the first disagreement between theory and experiment. Whether this disagreement was due to the simplifying approximations made in the model, the uncertainties in physical constants, or the details of the physical processes included in the model was investigated. By numerically solving the equations for several valves of the parameters in the theory, Latham and Smirl generated a large library of figures in which theory and experiment were compared. The culmination of this theoretical study was the paper by W. P. Latham, A. L. Smirl, A. Elci, and J. S. Bessey[4], which is presented as Appendix D of the report. Finally, a review paper, in which diffusion is included in the model for the first

time, was presented by A. Elci, A. L. Smirl, M. O. Scully, and C. Y. Leung.[5] A copy of this paper is also included as Appendix B of this report.

In summary, throughout the year, this project provided the theoretical and numerical support needed to compare theory with experiment. Members of this project submitted several publications. The initial paper by Elci, et al.[1] has been published in Physical Review B. The paper by Bessey, et al.[3] has been submitted for publication. Both the paper by Latham, et al.[4] and Elci, et al.[5] were presented at the International Conference on Hot Electrons in Semiconductors, which was held at North Texas State University, and have been accepted for publication by Solid State Electronics. All of the members of this project actively participated in the conference. A. L. Smirl, principal investigator, served as conference vice-chairman.

Experiment

In the first year of this grant, the experimental effort has been devoted to designing, constructing, and assembling a system that enables us to directly monitor ultrafast processes. This system utilizes an excitation-probe technique first used by Shelton and Armstrong[6] to monitor the evolution of nonequilibrium processes on a picosecond time scale.

This excitation-probe technique requires the use of the following equipment (see Figure 1):

(1) A mode-locked laser producing trains of ultrashort pulses.

(2) An electro-optic shutter to separate one of the ultrashort pulses from the train of pulses produced by the mode-locked laser.

(3) A detection system to monitor the pulse train and the shutter.

(4) An apparatus to split the "switched-out" pulse and temporally delay one part with respect to the other.

(5) An optical dewar to mount the sample and to control its temperature.

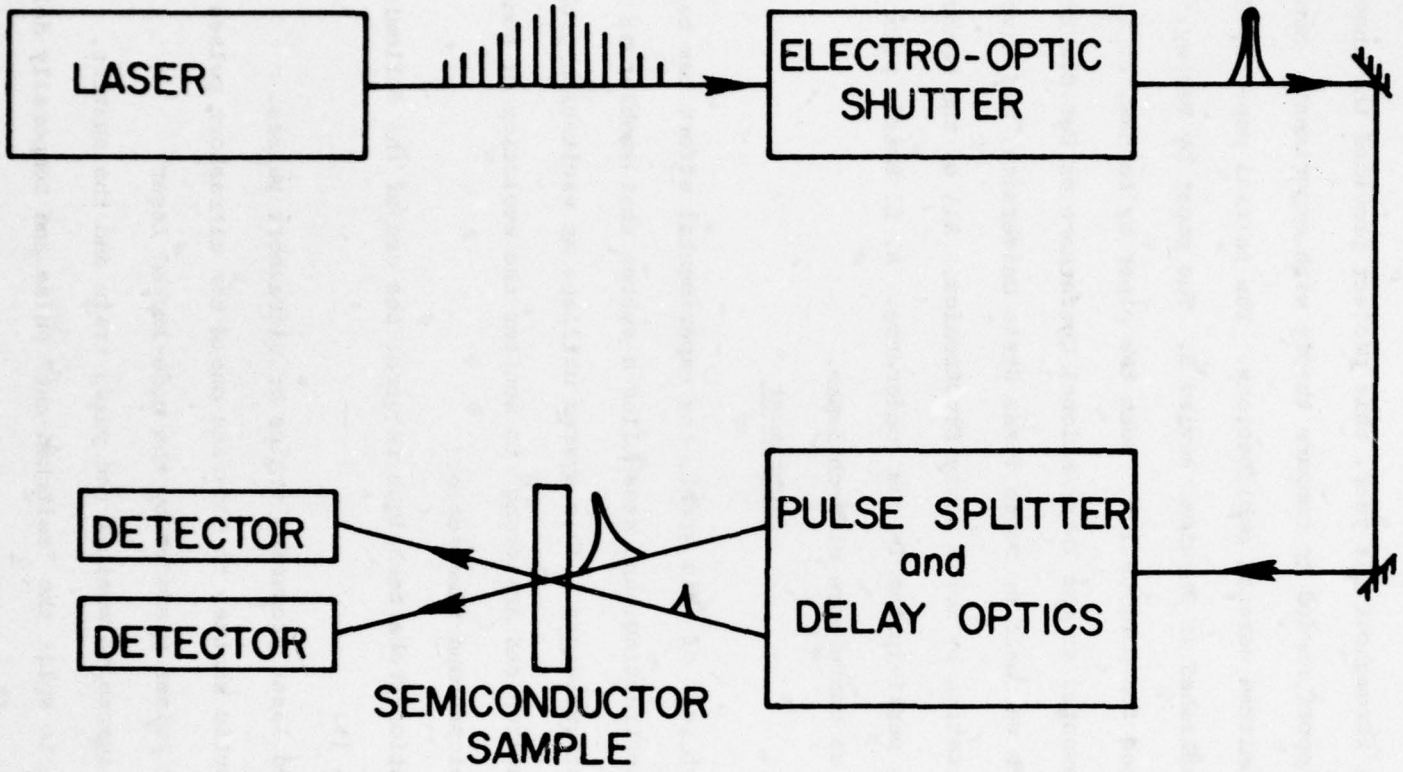


Fig. 1. Block Diagram of Experimental Arrangement for Measuring Ultrashort Relaxation Times in Semiconductors.

(6) A detection system to monitor the "switched-out" pulse both before and after the sample.

This last year has been spent in designing, constructing, assembling, and testing every component of the system described above as well as other pieces of peripheral equipment.

The mode-locked laser system consists of an end mirror, a laser head containing a xenon flashlamp, a Nd:glass rod, a cooling system, an output coupler, a dye cell in contact with the output coupler, a power supply for the flashlamp, and an alignment laser to ensure the alignment of the laser cavity.

This was the first part of the system to be assembled and tested. Mode-locking the laser requires that the dye concentration be balanced correctly with the power applied to the xenon flashlamp to ensure that the rod lases only once during the symmetrically shaped flashlamp pulse. As a consequence the dye concentration (i.e. the volume to volume ratios of 1-2 dichloroethane and an Eastman Kodak dye, A9860) was varied from 10:1 to 3:1 while the laser output was monitored at each concentration as a function of voltage applied to the flashlamp. Two monitoring processes were used. The output of the laser was monitored directly using a fast risetime biplanar photodiode-oscilloscope combination. Also the output was monitored indirectly using a silicon PIN-photodiode in the photoconductive mode with an oscilloscope, and the laser radiation was observed as a spike riding on top of the symmetric flashlamp pulse. As a result of these measurements a 5:1 dye solution was chosen as the best match to our needs. It is necessary to point out, however, that mode-locking is such a complicated statistical process that we have from time to time run into difficulties in obtaining clearly mode-locked output from the laser system. At such a time, one must check all of the

possible causes of problems (e.g. bad dye, dirty dye cell) until the problem is corrected.

The next piece of equipment that was installed was the electro-optic shutter (see Fig. 2) consisting of two Glan-Foucault polarizers, a Pockel's cell, a spark gap, a d.c. power supply, and coaxial cable to connect the power supply, spark gap, and Pockel's cell. Among the important problems associated with the set-up of the shutter were:

(1) One would like the ratio of light passed by the shutter with voltage applied to the Pockel's cell compared with the light passed by the shutter with no voltage on the cell to be as large as possible. In order to maximize this ratio (the switch-out ratio) the polarizers must be crossed precisely before the Pockel's cell is placed between them. The switch-out ratio was tested by monitoring the light passed with no voltage applied to the Pockel's cell and comparing it to the light passed when a single pulse was switched from the pulse train by the shutter. This ratio was measured to be approximately 100:1 and should not be confused with the usual extinction ratio.

(2) The KD*P crystal in the Pockel's cell must be aligned very precisely with its optical axis parallel to the direction of propagation of the pulse train. Alignment of the Pockel's cell was ensured by viewing the crystal between the crossed polarizers while it was illuminated by an extended, diffuse source and by orienting the crystal in such a manner that a dark or opaque area shaped like a Maltese cross appeared centered in the aperture of the polarizers.

(3) Breakdown of the spark gap must be controlled in one of two ways: either by adjusting the gap spacing or by adjusting the gas pressure. The product of gas pressure and gap spacing must be chosen so that breakdown

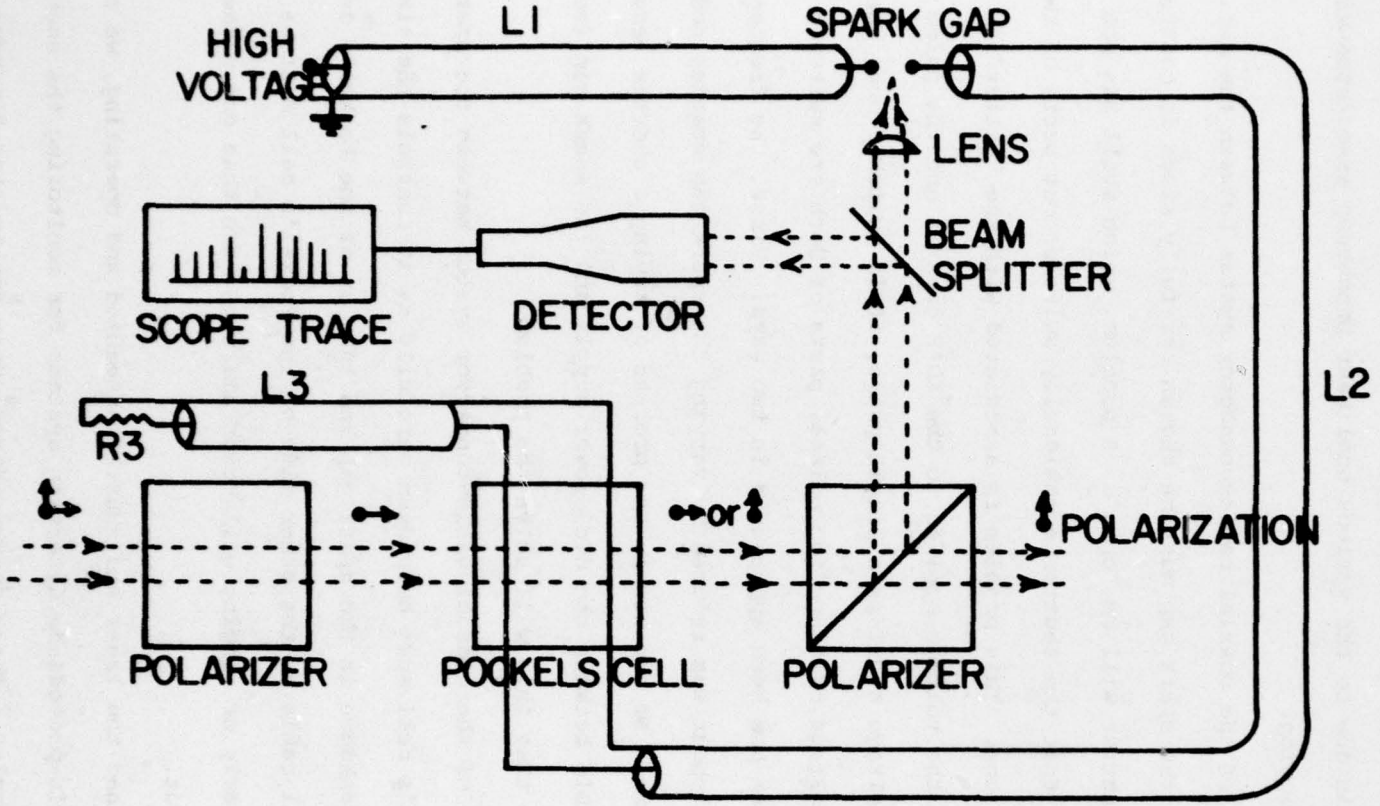


Fig. 2. Schematic Diagram of the Electro-Optic Shutter System.

occurs early in the pulse train. One of the first few pulses must be switched from the pulse train since the pulsewidth increases at the back of the train due to the various nonlinear processes associated with the pulse production.

(4) The length of the coaxial cable-connector system between the d.c. power supply and the spark gap must be chosen carefully since it controls how long the shutter will be "open". A problem, which still has not been solved, is that the shutter occasionally switches out parts of two pulses instead of one. This problem is associated with the "jitter" in the overlap of the voltage applied to the KD*P crystal and the pulse train. When the voltage overlaps a single pulse, it is switched out; whereas, when the voltage overlaps two pulses, parts of both are switched out. This problem has been approached in two ways. First, the frequency of double switchout was reduced by varying the spark gap spacing and the gas pressure. We are now in the process of trying a shorter length of coaxial cable between the d.c. power supply and the spark gap (see Fig. 2). We believe that this will solve the problem.

(5) The length of the coaxial cable-connector system between the spark gap and the Pockel's cell must be chosen carefully as it controls the time delay between breakdown in the spark gap and turn on of the Pockel's cell.

(6) The coaxial cable on the other side of the Pockel's cell must be terminated properly or ringing will occur allowing more than one pulse to be switched out.

Once we had the laser and shutter assembled and operating, we calibrated the silicon PIN-photodiode detection systems for monitoring the energy of the switched-out pulse. These energy meters[4] were constructed "in house" at a

cost of approximately $50.00 apiece. The detectors were checked for linearity and found to linear over a 1-9 volt range. The absolute calibration of the detectors was obtained by comparison of their readings with those obtained from a Scientech Laser Energy Meter and sensitivities were obtained for each of the detectors.

In addition, two detector systems that monitor the pulse trains and the electro-optic shutter have been installed and tested. The first consisted of a Hamamatsu R617U biplanar photodiode in combination with a Tektronix 7904 oscilloscope with a fast write option. This combination (mainly the oscilloscope) proved to be less than optimal for our purposes and the oscilloscope is presently being used on the other project funded by this grant. The second system consists of an ITT F4000 biplanar photodiode in combination with a Tektronix 519 oscilloscope. This system works satisfactorily although we are now redesigning the housing on the biplanar photodiode to reduce its ringing.

The pulse splitting and delay optics (see Fig. 3) consist of a beam splitter, a stationary 90° prism (mounted on a very accurate Klinger Scientific translation stage providing a resolution of 0.01 mm and an accuracy of 3 mm over the 40 mm displacement allowed) and two mirrors on orthoganally adjustable mounts. Two of the problems associated with this piece of equipment are:

(1) The paths traveled by the different pulses separated by the beamsplitter must be measured precisely so that the path containing the translation stage can be adjusted from a few picoseconds shorter than the fixed path to many picoseconds longer than the fixed path. This can be done only coarsely by mechanical measurement techniques, and zero delay must be determined by more accurate optical techniques.

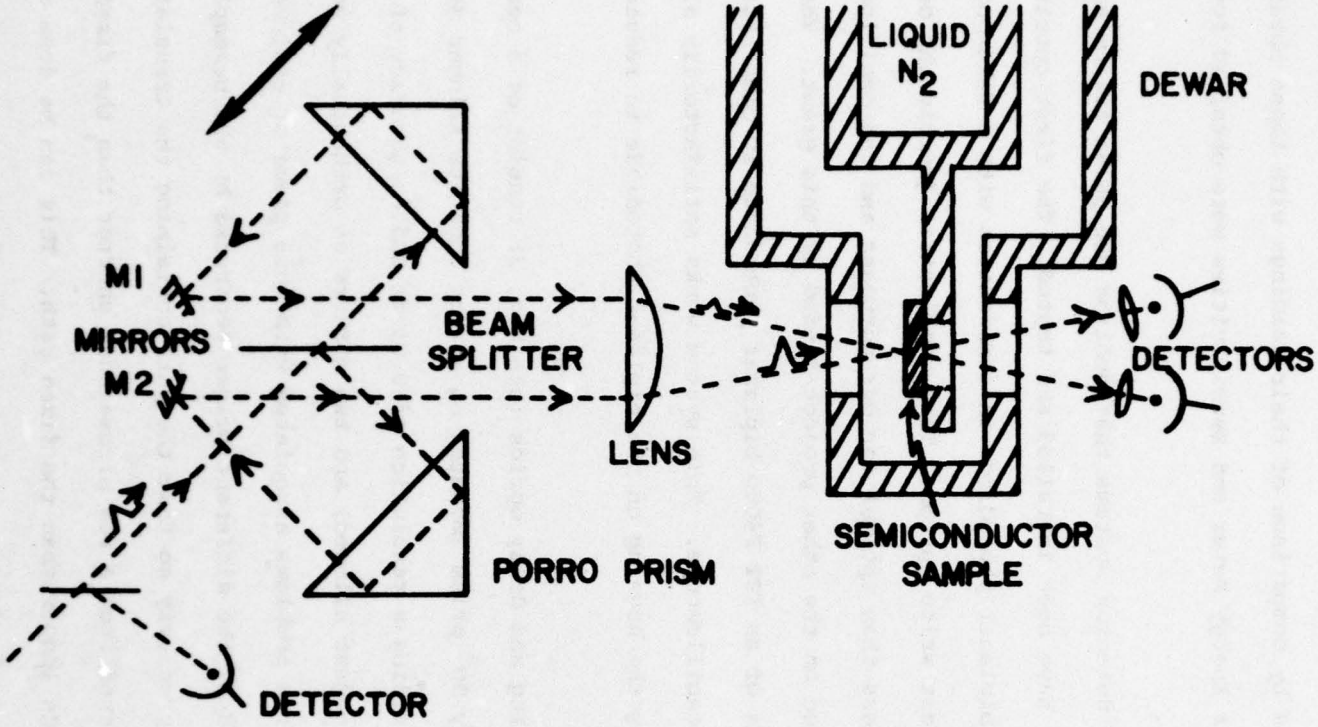


Fig. 3. Schematic Diagram of Pulse Splitter and Delay Optics

(2) The two pulses must overlap at the sample surface requiring tests to be made before and after each experiment utilizing burn patterns on a piece of photographic emulsion positional in the plane of the sample.

The optical dewar itself consists of a relatively inexpensive brass container with cold finger and a brass jacket with four optical ports covered by 3 inch quartz optical flats. The dewar is evacuated by a "roughing" pump and is capable of lowering the sample temperature to approximately 100 K. The dewar sits on a NRC lab jack and is capable of being moved vertically through 3 inches.

Finally, the latest piece of equipment to be installed and tested is a flowing dye cell for the mode-locked laser. The old arrangement consisted of a dye reservoir, a dye cell, and a hypodermic syringe all connected by teflon tubing. Every few laser firings new dye had to be pulled into the dye cell by adjusting the syringe. The new flowing system consists of the dye reservoir, dye cell, and a peristaltic pump which pulls the dye solution through the cell at a rate of 0.2 cc/minute.

Once the entire system was assembled and each of the individual components had been shown to be functioning properly, it was necessary to test the system as a unit. This was done by repeating the single-pulse and excitation-probe experiments performed on germanium by Kennedy, et al.[7,8] and Shank and Auston[9]. These experiments have been completed and the results are shown in Fig. 4 and Fig. 5, and the results are in good agreement with previous experiments.

In conclusion, experimentally, the first year of this grant has been devoted to developing and ensuring the operational integrity of the basic system. Each component of the system has been installed and tested rigorously. The system as a whole has been tested by comparing data obtained using the present system with data obtained at other laboratories.

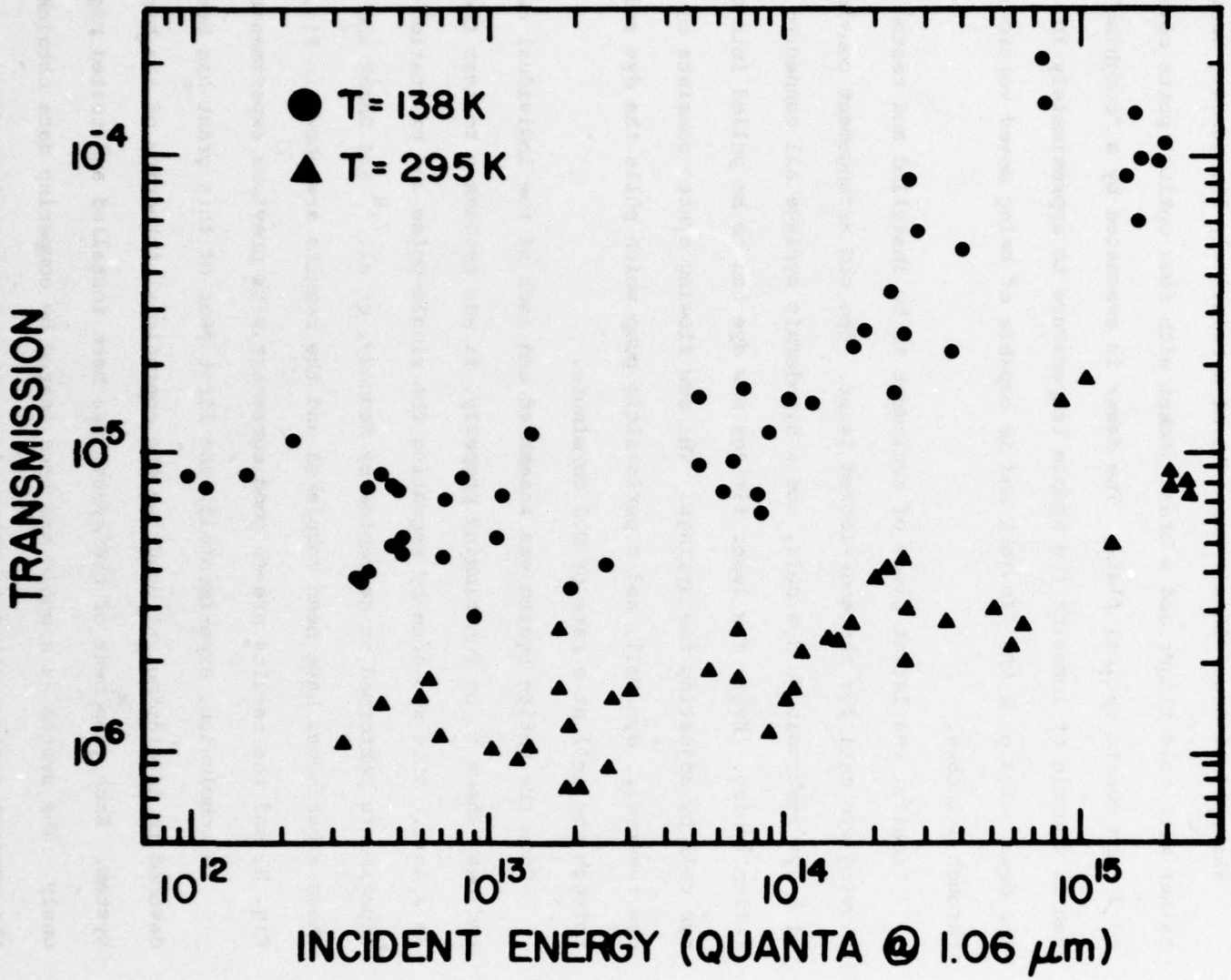


Fig. 4. Single Pulse Transmission of a Single Crystal Germanium Sample Versus Incident Energy in Units of Quanta at 1.06 μm.

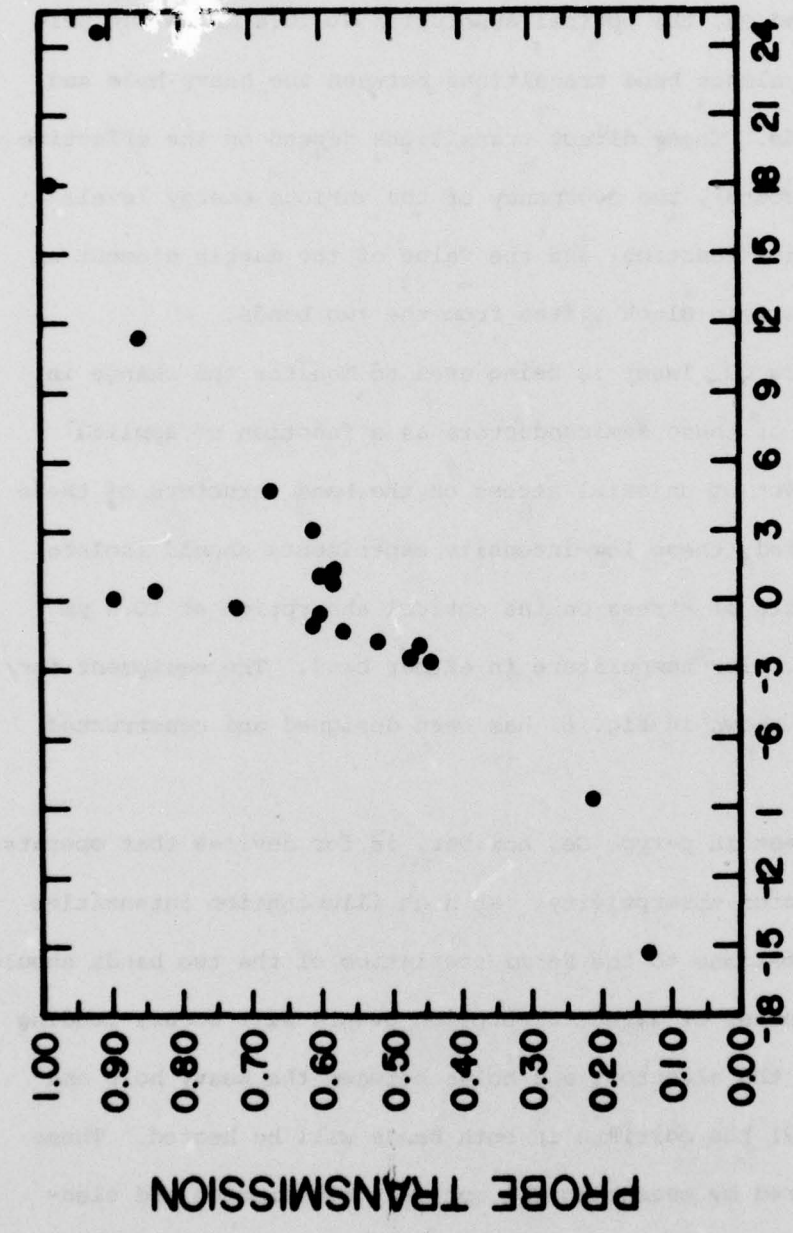


Fig. 5. Normalized Ratio of Probe Pulse Transmission to Excitation Pulse Transmission Versus Time Delay of Probe Pulse with Respect to Excitation Pulse at Temperature 297 K.

STRESS AND ELECTRIC FIELD BIASING OF OPTICAL PROPERTIES OF SEMICONDUCTORS

For p-type GaAs and Ge, the optical absorption at 10.6 μm is due primarily to direct intervalence band transitions between the heavy hole and light hole valence bands. These direct transitions depend on the effective mass (energy band structure), the occupancy of the various energy levels (Fermi-Dirac distribution function) and the value of the matrix element of the crystal momentum between Block states from the two bands.

A low-intensity, cw CO_2 laser is being used to monitor the change in the optical absorption of these semiconductors as a function of applied stress. Since the effect of uniaxial stress on the band structure of these materials is well studied, these low-intensity experiments should isolate and calibrate the effects of stress on the optical absorption at 10.6 μm without altering the carrier temperature in either band. The equipment for these measurements, as shown in Fig. 6, has been designed and constructed in-house.

Much of the interest in p-type Ge, however, is for devices that operate in the region of saturated absorptivity. At high illumination intensities two significant perturbations to the Fermi statistics of the two bands should result: (1) A large number of direct absorption events will occur, leading to a redistribution of the electrons and holes between the heavy hole and light hole bands and (2) the carriers in both bands will be heated. These changes will be monitored by measuring the optical transmission and electrical conductivity. In order to provide high intensity illumination of

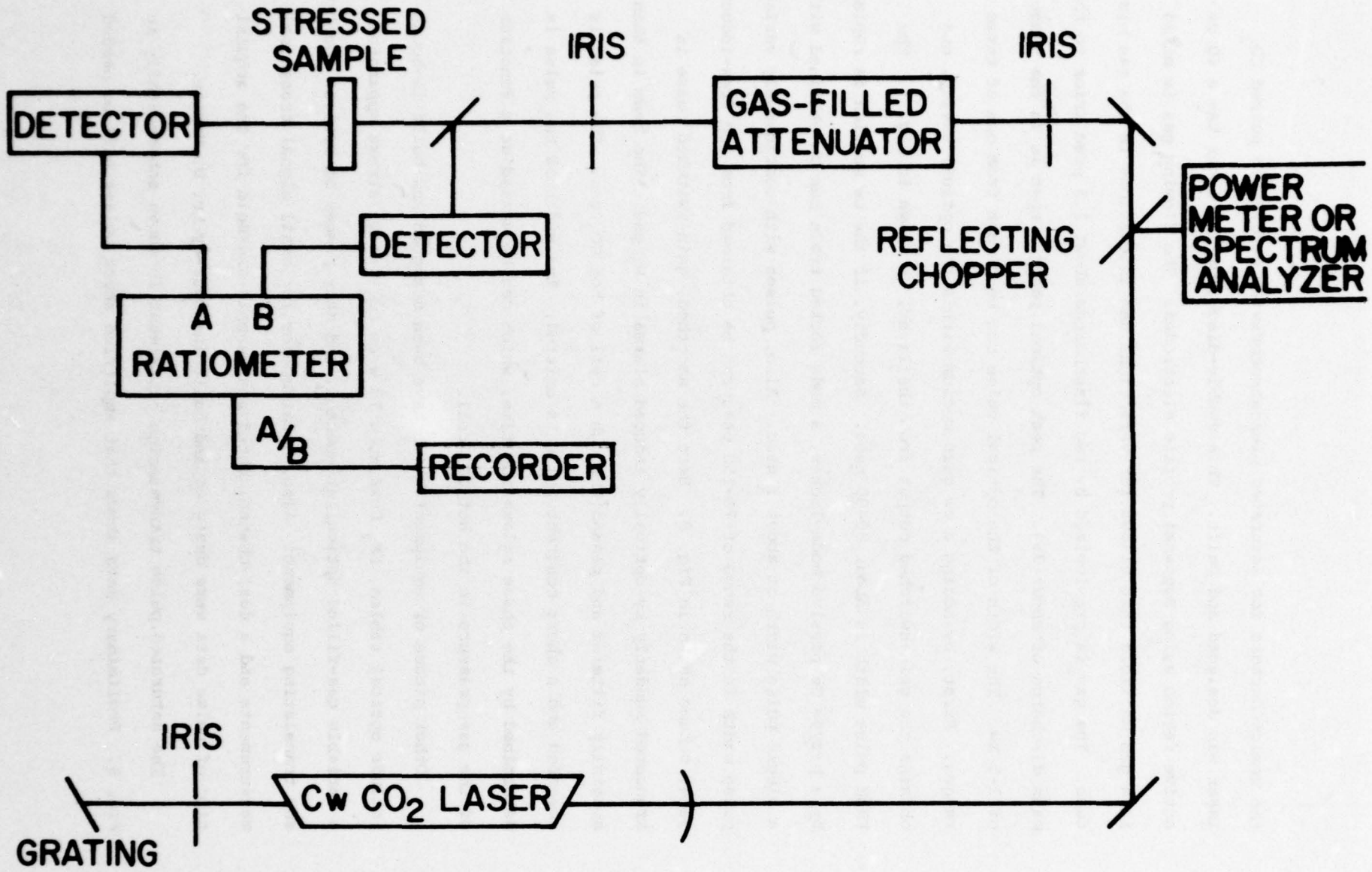


Fig.6. Small signal transmission scheme.

the semiconductors for saturated piezoabsorptivity studies, a pulsed CO_2 laser was designed and built. This double-discharge TEA laser has a 60 cm active region using Rogowski profile electrodes. The flowing gas is mixed in a gas manifold that allows for variations and innovations in the gas mixture. The gas is pre-ionized by two flashboards about 1.5 μsec prior to the main discharge of about 35j. The peak optical power output is in the range of 1-5 Mw. The width of the optical pulse can be chosen from one of three ranges. First, by adding a cw gain medium within the optical cavity, one obtains the gain-switched output from the laser, as shown in Fig. 7. The FWHM pulse width is about 80-90 nsec. Secondly, if the cw section is replaced by a p-type Ge passive mode-locker, a mode locked train can be obtained with a single pulse width of about 1 nsec. Also, pulses with continuously variable pulse width in the range of 70-150 psec can be obtained from the free-induction decay scheme shown in Fig. 8. Here the smoothed, gain-switched pulse is truncated suddenly by optically induced plasma in N_2 gas. The beam is then spatially filtered and passed through a cell of hot CO_2 gas. The pulse is absorbed and a short coherent pulse is emitted. The width of the pulse is determined by the phase relaxation time, which can be varied as a function of the gas pressure in the hot CO_2 cell.

Other pieces of equipment which have been designed and built in-house include optical tables, LN_2 Dewars, a 10 w cw CO_2 laser, stress apparatus, a variable gas-filled optical attenuator, and many pieces of optical mounting and translating equipment. Also, a ratiometer for small signal transmission measurements and a dual-channel, gated peak-sample-and-hold for the acquisition of pulse data were designed and built in-house by Mr. D. Maxson.

The saturated pulse transmission experiment is shown schematically in Fig. 9. Preliminary data shows that amplified short pulses will be needed

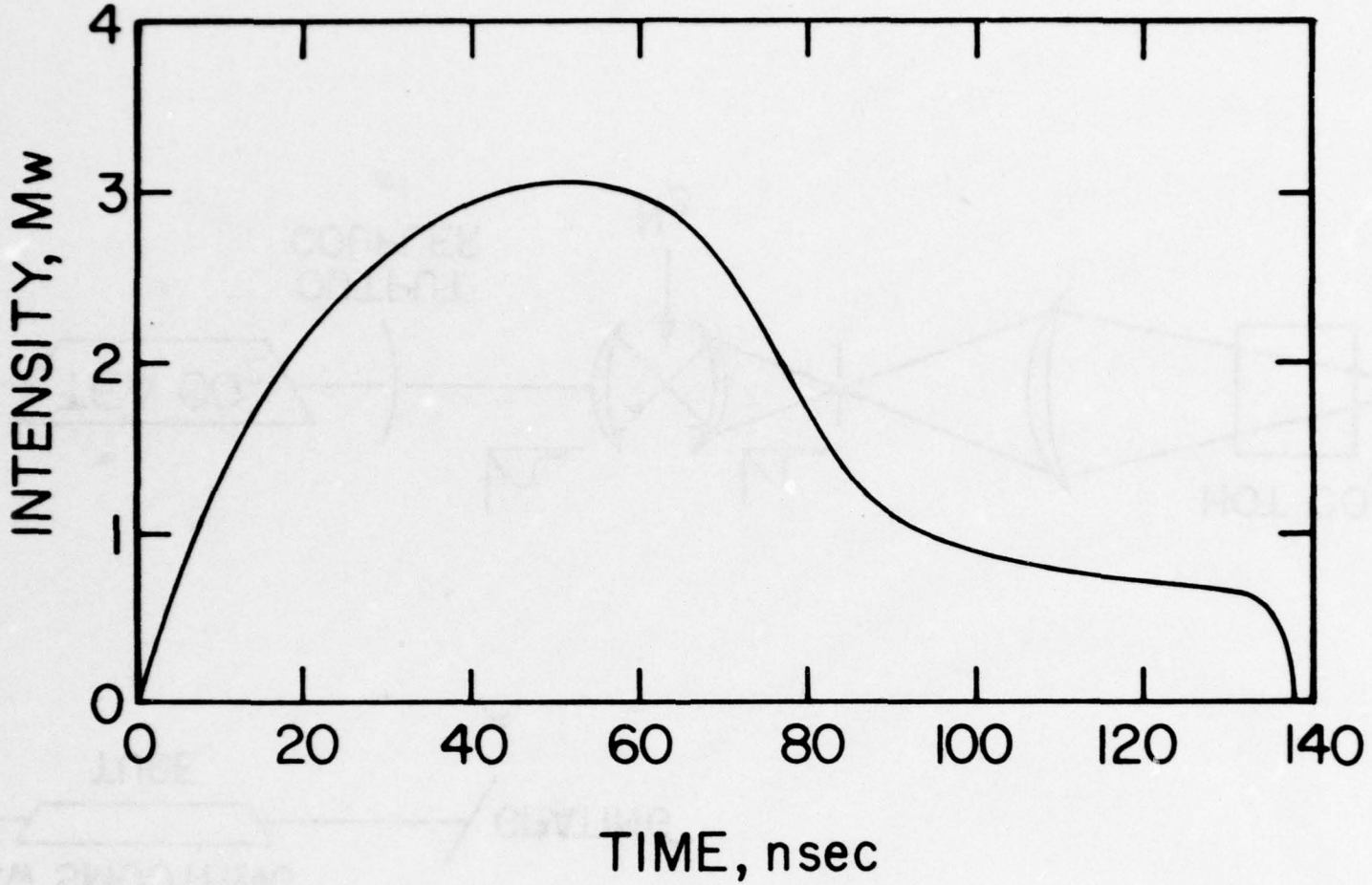


Fig. 7. Typical smoothed gain-switched CO_2 pulse.

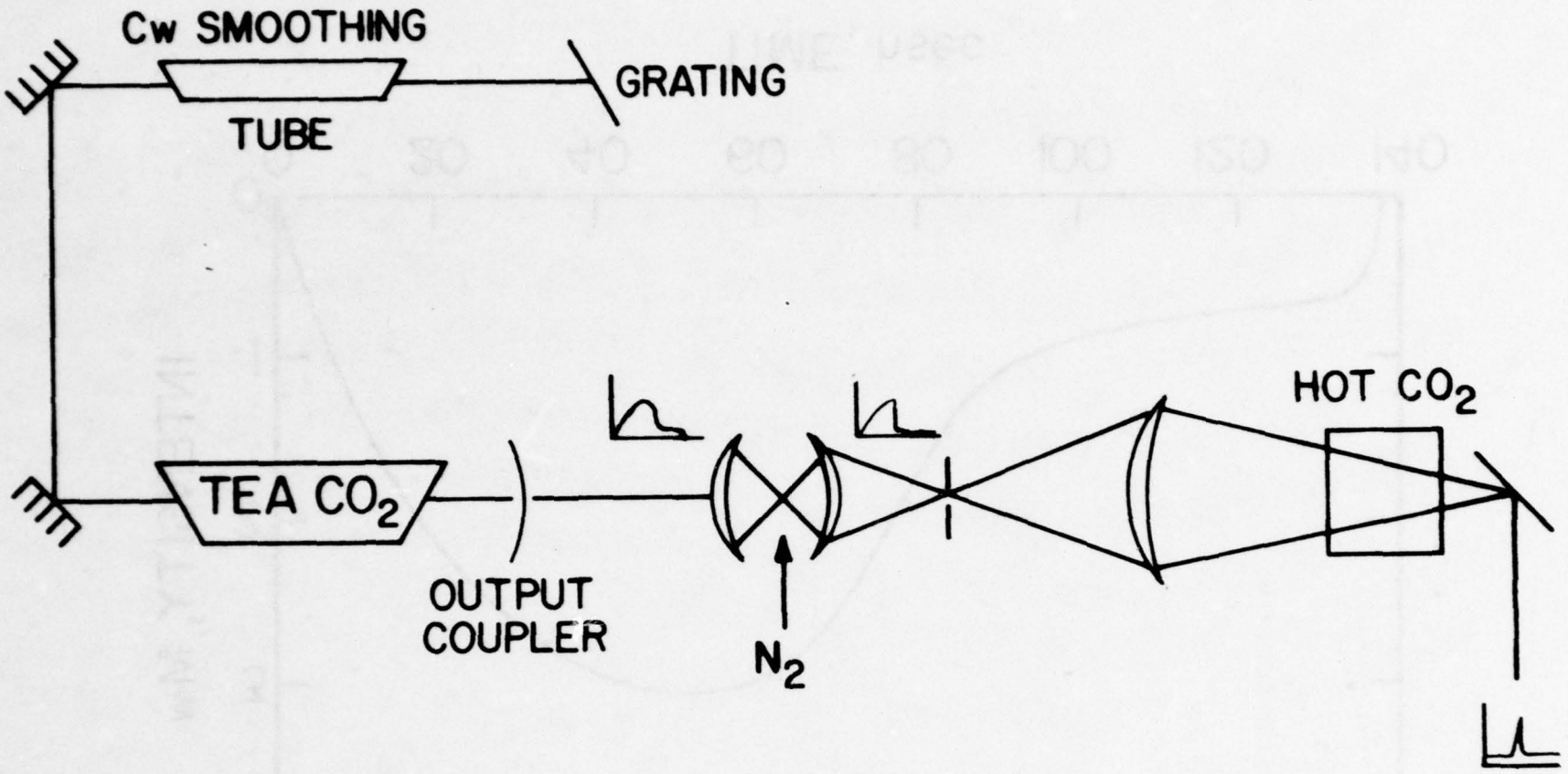


Fig. 8. Short pulse generation by free-induction decay.

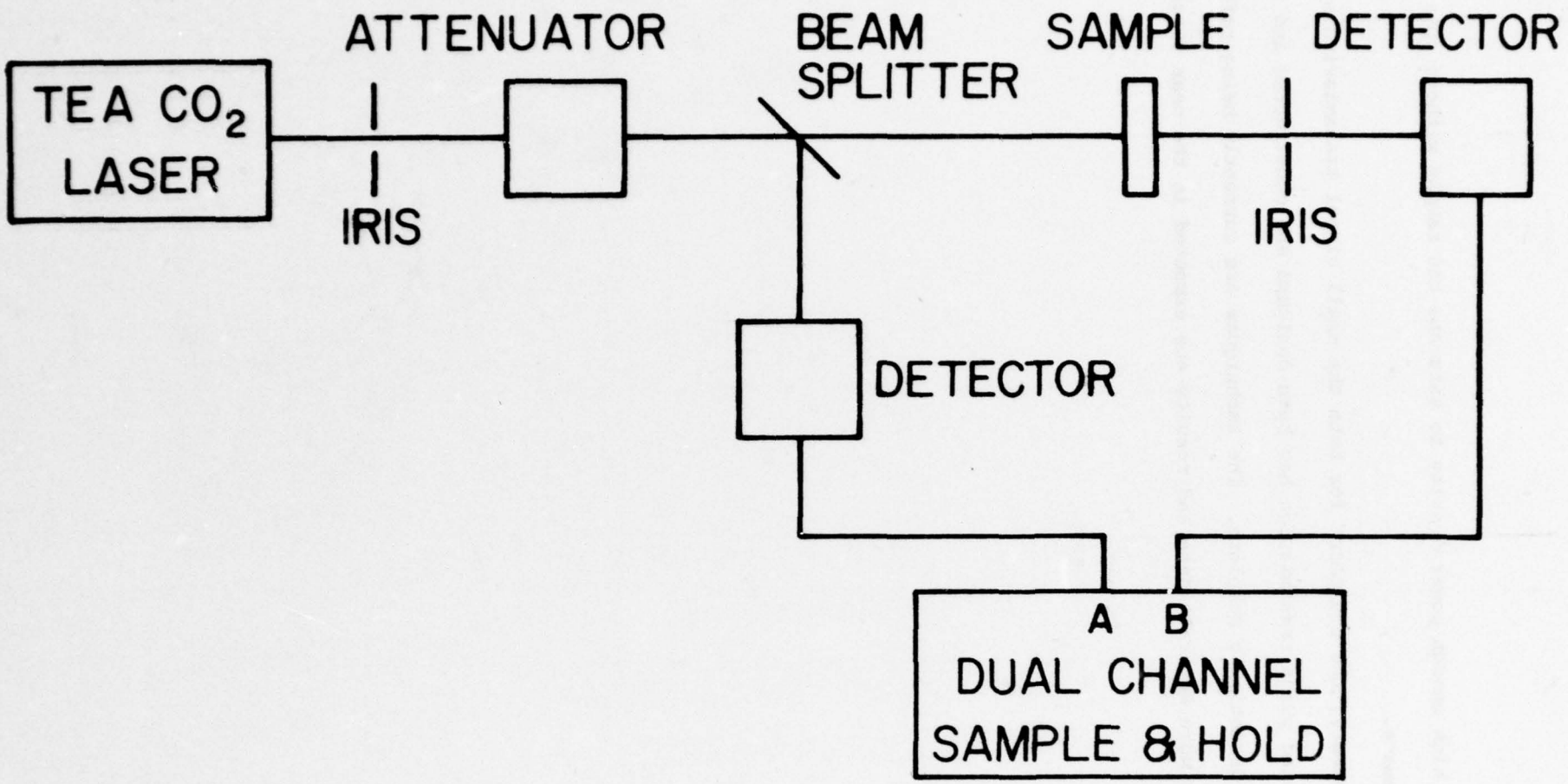


Fig. 9. Saturated pulse transmission experiment.

to obtain high enough power density to saturate the sample without optical surface damage.

In summary, the equipment for both the small signal transmission and the saturated pulse transmission has been designed and constructed and the system is essentially complete. The techniques are currently being refined to produce more accurate data and results are expected in the near future.

PROFESSIONAL PUBLICATIONS AND ACTIVITIES

FOR THE PERIOD

1 SEPT. 1976 to 31 AUG. 1977

Arthur L. Smirl - Principal Investigator

Publications

1. "Ultrafast Transient Response of Solid State Plasmas: I. Germanium, Theory and Experiment," Ahmet Elci, M. O. Scully, A. L. Smirl, and J. C. Matter, Phys. Rev. B1 16, 191 (1977).

2. "On the Pulsewidth Dependence of the Transmission of Ultrashort Optical Pulses in Germanium," J. S. Bessey, B. Bosacchi, H. M. van Driel, and A. L. Smirl, submitted to Phys. Rev. B, (1977).

3. "The Role of Phonons and Plasmons in Describing the Pulsewidth Dependence of the Transmission of Ultrashort Optical Pulses through Germanium," W. P. Latham, Jr., A. L. Smirl, A. Elci, and J. S. Bessey, accepted for publication by Solid-State Electron., 1977.

4. "Physics of Ultrafast Phenomena in Solid State Plasmas," A. Elci, A. L. Smirl, C. Y. Leung, and M. O. Scully, accepted for publication by Solid-State Electron., 1977.

5. "Gauge Invariant Perturbation Theory for the Interaction of Electromagnetic Radiation and Matter," D. H. Kobe and A. L. Smirl, submitted to Am. J. Phys., 1977.

Papers

1. "Ultrafast Transient Response of Optically Excited Plasmas in Germanium," Invited paper presented at 1976 Annual Meeting of Optical Society of America, J. Opt. Soc. Am. 66, 1082 (1976).

2. "The Role of Phonons and Plasmons in Describing the Pulsewidth Dependence of the Transmission of Ultrashort Optical Pulses Through Germanium," A. L. Smirl, W. P. Latham, A. Elci, and J. S. Bessey, International Conference on Hot Electrons in Semiconductors, Bull. Am. Phys. Soc. 22, 706, (1977).

3. "Gauge Invariant Formulation of the Interaction of Radiation and Matter," D. H. Kobe and A. L. Smirl, AAPT Summer Meeting, 1977, AAPT Announcer 7, 70 (1977).

Other Talks

1. "Ultrashort Optical Pulse Measurement at 10.6 Microns," Larry Tipton A. L. Smirl, and D. G. Seiler, Texas Academy of Science Meeting at Baylor University, March 10, 1977.

2. "Ultrafast Transient Response of Hot Carriers in Germanium," S. C. Moss, J. R. Lindle, A. L. Smirl, Texas Academy of Science Meeting at Baylor University, March 10, 1977.

3. "Gauge Transformations and Perturbation Theory," D. H. Kobe and A. L. Smirl, Texas Academy of Science Meeting of Baylor University, March 10, 1977.

David G. Seiler - Faculty Associate

Publications

1. "The Shubnikov-de Hass Effect in n-InAs and n-GaSb" (with A. E. Stephens, R. E. Miller, and J. R. Sybert), to be published in Phys. Rev. B 15.

2. "Inversion-Asymmetry Splitting of the Conduction Band in InSb" (with B. D. Bajaj and A. E. Stephens), to be published in Phys. Rev. B 15.

3. "A Magnetic Field Modulation Technique for the Study of Hot Carrier Oscillatory Magnetoresistance Phenomena" (with H. Kahlert), to be published in Rev. Sci. Instrum.

4. "Electric-Field Dependence of the Positions and Amplitudes of Magnetophonon Oscillations in n-InSb at 77K" (with H. Kahlert and J. R. Barker), to be published in Solid State Electronics.

5. "CO_2 Laser-Induced Hot Electron Effects in n-InSb" (with B. T. Moore and H. Kahlert), to be published in Solid State Electronics.

6. "Observations of Magnetophonon Structure in Degenerate n-InSb" (with H. Kahlert), to be published in Solid State Communications.

Papers

1. "High Resolution Measurements of the Hot-Electron Magnetophonon Effect in n-InSb at 77k" (with H. Kahlert and A. E. Stephens), Bull. Am. Phys. Soc. 22, 460 (1977).

2. "Shubnikov-de Haas Effect in n-InAs and n-GaSb" (with A. E. Stephens, R. E. Miller, and J. R. Sybert), Bull. Am. Phys. Soc. 22, 460 (1977).

3. "Laser-Induced Hot Electron Transport Effects in n-InSb at 2K" (with B. T. Moore and H. Kahlert), Bull. Am. Phys. Soc. 22, 460 (1977).

4. "A Magnetic Field Modulation Technique for the Study of Hot Carrier Oscillatory Magnetoresistance Phenomena" (with H. Kahlert), Bull. Am. Phys. Soc. 22, 460 (1977).

5. "Electric-Field Dependence of the Positions and Amplitudes of Magneto-phonon Oscillations in n-InSb at 77K) (with H. Kahlert and J. R. Barker), Bull. Amer. Phys. Soc. 22, 708 (1977).

6. "CO_2 Laser-Induced Hot Electron Effects in n-InSb" (with B. T. Moore and H. Kahlert), Bull. Amer. Phys. Soc. 22, 709 (1977).

REFERENCES

1. A. Elci, M. O. Scully, A. L. Smirl, and J. C. Matter, Phys. Rev. B 16, 191 (1977).

2. H. M. van Driel, J. S. Bessey, and R. C. Hanson, to be published.

3. J. S. Bessey, B. Bossachi, H. M. van Driel, and A. L. Smirl, submitted for publication.

4. W. P. Latham, Jr., A. L. Smirl, A. Elci, and J. S. Bessey, accepted for publication by Solid State Elec.

5. A. Elci, A. L. Smirl, M. O. Scully, and C. Y. Leung, accepted for publication by Solid State Elec.

6. J. W. Shelton, and J. A. Armstrong, IEEE JQE QE-3 696 (1967).

7. C. J. Kennedy, J. C. Matter, A. L. Smirl, H. Weichel, F. A. Hopf, and S. V. Pappu, Phys. Rev. Letters 32, 419 (1974).

8. A. L. Smirl, J. C. Matter, A. Elci, and M. O. Scully, Opt. Commun. 16, 118 (1967).

9. C. V. Shank, and D. H. Auston, Phys. Rev. Letters 34, 479 L1975).

APPENDIX A

ULTRAFAST TRANSIENT RESPONSE OF SOLID-STATE PLASMAS.

I. GERMANIUM, THEORY, AND EXPERIMENT

PHYSICAL REVIEW B VOLUME 16, NUMBER 1 1 JULY 1977

Ultrafast transient response of solid-state plasmas. I. Germanium, theory, and experiment*

Ahmet Elci and Marlan O. Scully
Department of Physics and Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Arthur L. Smirl
Department of Physics, North Texas State University, Denton, Texas 76203

John C. Matter
Sandia Laboratories, Albuquerque, New Mexico 87115

(Received 7 June 1976)

A first-principles theory is developed for the generation and the subsequent transient behavior of dense electron-hole plasmas produced in germanium by intense picosecond optical pulses. Experimental data are discussed and compared with the theory. It is shown that the valley structure of the Ge conduction band, phonons, and plasmons play significant roles in the generation and the temporal evolution of the plasma. The agreement between the theory and experiments is good. Some predictions of the theory are discussed.

I. INTRODUCTION

The development of mode-locked lasers[1] that produce optical pulses of a few picoseconds duration and peak powers of $\gtrsim 10^9$ W has made possible recent experimental investigations of ultrafast electronic processes in semiconductors.[2-5] In these experiments, the general method is to subject an intrinsic semiconductor first to an intense picosecond (excitation) pulse[2] of frequency ω_0 such that $\hbar\omega_0$ is greater than the band-gap energy of the material and to measure its transmission; then after some delay, to subject the same sample to a second (probe) pulse of the same frequency but lower intensity and measure its transmission. The transmission properties of the semiconductor are significantly altered when the excitation pulse creates a large free-carrier density, and the transmission of the probe pulse yields information concerning the temporal evolution of the free-carrier distribution. The ultrashort duration of the laser pulses makes it possible to follow this temporal evolution on a picosecond time scale.

In this paper we present a first-principles, theoretical treatment accounting for the observed behavior in the excitation and probe pulse measurements in terms of the energy-band structure and other well-known parameters of germanium (Ge). We note at the outset, however, that only certain of the myriad of possible electronic interactions are included in the present work. Future papers in this series will deal with other possible electron processes such as diffusion and Kane-like excitations.

In the remainder of this introductory section, we briefly review the experimental method and results (Sec. I A), summarize the pertinent germanium energy-band structure (Sec. I B), and qualitatively discuss the physical processes which occur during and after the nonlinear light absorption by germanium (Sec. I C). We expound on the theory in Sec. II. This elaboration proceeds through a discussion of electron-hole distributions (Sec. II A), direct absorption (Sec. II B), free-carrier absorption (Sec. II C), phonon-assisted relaxation (Sec. II D), plasmon-assisted recombination (Sec. II E), rate equations for absorption and transmission (Sec. II F), and integration of the dynamical equations (Sec. II G). The final section (Sec. III) of the main text contains the comparison of theory with experiment and our conclusions. Several appendixes include detailed calculations which allow us to make critical simplifying assumptions in our theoretical model. Appendix A considers intervalley transitions via phonon emission; Appendix B discusses Coulomb thermalization; Appendix C derives radiation absorption rates; Appendix D expands on phonon-assisted relaxation. Finally, in Appendix E, we present some details concerning the experimental method and apparatus.

A. Experimental method and data

A mode-locked Nd:glass ($\hbar\omega_0 = 1.17$ eV) laser produces a train of pulses; each pulse is typically 5–10 psec in duration and has $\sim 10^{15}$ quanta. Using a laser-triggered spark gap and electro-optical shutter, a single pulse can be isolated from this train of pulses.[6]

In the first experiment, this single pulse is passed through a variable attenuator and focused on an intrinsic 5.2-μm-thick Ge sample [Fig. 1(a)]. The focused spot is approximately 250 μm in diameter. The transmission of this pulse is then mon-

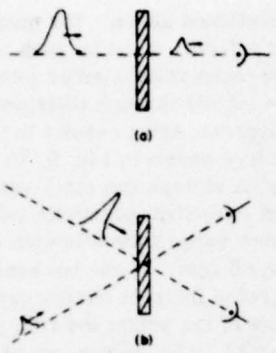


FIG. 1. Schematic diagram of picosecond experiments for (a) saturation and (b) excitation pulse/probe studies.

itored. The transmission data as a function of incident number of quanta, for two initial sample temperatures (77 and 297 °K), are shown in Fig. 2. At low pulse energies, transmission follows Beer's law.[7] As the incident number of quanta increases, transmission increases and eventually saturates. Near 5×10^{14} quanta samples are damaged.

In the second experiment [Fig. 1(b)], the sample is first irradiated by an excitation pulse of sufficient energy to cause enhanced transmission. It is then followed, at a later time t_d, by a probe pulse which is about 5% as intense as the excitation pulse. The excitation and probe pulses are derived from a single pulse as described in Appendix E. Probe pulse transmission data versus delay time t_d are shown in Fig. 3 for initial sample temperatures of 77 and 297 °K and for a fixed excitation pulse of about 10^{14} quanta. There is a sharp increase in transmission, a spike, in the region of temporal coincidence of the probe and excitation pulses (at $t_d = 0$). The spike is due to the scattering of the excitation pulse into the probe pulse by an index grating formed by the interference of these two coherent pulses in the crystal.[4,8] In this paper,

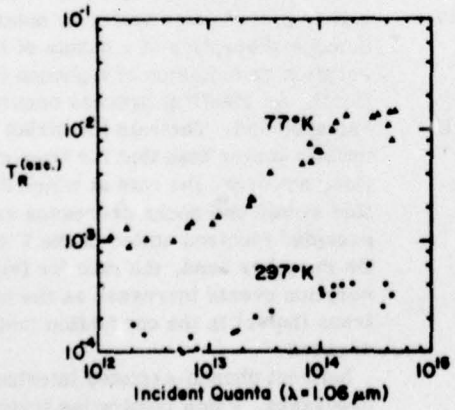


FIG. 2. Excitation pulse transmission (saturation) data.

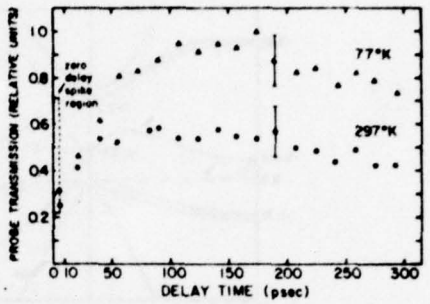


FIG. 3. Probe pulse transmission data. Relative transmission units are given by the ratios of the excitation transmission to probe transmission for each delay, normalized such that the peak of the 77 °K curve is unity.

we will be concerned with the probe transmissions at delay times longer than the pulse width. In this region, probe transmission increases relatively rapidly, reaching a maximum near $t_d \sim 100$ psec, thereafter decreasing slowly. Figure 4 shows the probe pulse data for different excitation pulse intensities and for an initial sample temperature of 77 °K. Ratios labeling these curves are the normalized relative transmissions of excitation pulse to probe pulse. The details of the experiments are further discussed in Appendix E.

B. Review of Ge energy structure

The energy-band structure of Ge is well known,[9] and is shown in Fig. 5. The significant features of the conduction band are the locations of the conduction-band valleys. The minima of these valleys are quite close in energy. The minimum located at Γ is separated from the top of the valence band by 0.805 eV at 300 °K. This separation is 0.889 eV at 77 °K. The indirect gap located at L has a separation of 0.664 eV at 300 °K and a separation of 0.734 eV at 77 °K. The band minimum located near X is 0.18 eV higher than the minimum at L. There are three other valleys such as the one in the [111]

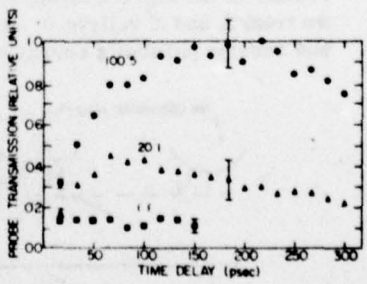


FIG. 4. Probe pulse transmission data for different excitation pulse intensities.

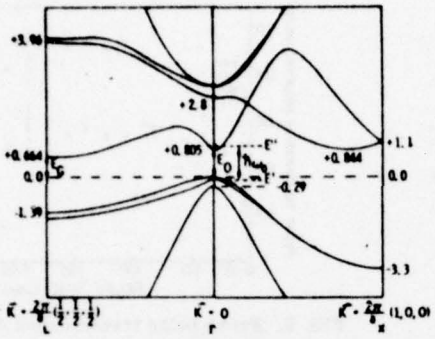


FIG. 5. Ge energy-band (eV) structure at 300 °K.

direction whose minima are located exactly at the Brillouin-zone boundary. There are five other valleys like the one in the [100] direction.[10] The constant energy surfaces in these valleys are elongated ellipsoids. The density-of-states effective mass[11] for the L valleys is $m_c \simeq 0.22m$.[9] The density-of-states effective mass of the central conduction-band valley is taken to be[9] $m_0 \simeq 0.1m$ at 300 °K and $0.04m$ at 77 °K.

The valence band maximum is at Γ. Two of the valence bands, which are degenerate throughout the Brillouin zone when spin-orbit coupling is neglected, we call heavy-hole bands, while the third is termed the light-hole band. At Γ, the light-hole band is lowered by spin-orbit coupling with respect to the heavy-hole bands by about 0.28 eV. Except for a small region near the center of the Brillouin zone, the structure introduced by the spin-orbit coupling in the heavy-hole bands is relatively minor. We will treat the two heavy-hole bands as degenerate. The effective mass of the two heavy-hole bands in the central region of the Brillouin zone is $m_h = 0.34m$.[9] We will ignore the light-hole band owing to its small contribution to the hole density of states and its large separation from the Γ conduction-band edge, which makes its contribution to light absorption negligible in the current experiments.

Let us remark that an enormous simplification results in the algebra of the subsequent sections if we treat L and X valleys in an equivalent manner and assume parabolic energy bands with the effective masses mentioned above. The quantitative results are not affected much by such an approximation because these valleys enter into the absorption properties mainly through their density of states. This approximation results in the simplified band structure shown in Fig. 6, in which the four "effective" L valleys and six X valleys are replaced by ten equivalent parabolic valleys. In each of these new valleys the effective mass for electrons is $m_c = 0.22m$. These ten equivalent valleys are at some indirect energy gap E_G, which is some average of the actual indirect gaps E_L (at L) and E_X (near X). The average we use is $E_G = \frac{1}{14}(8E_L + 6E_X)$. The new valleys are assumed to be centered in the Brillouin-zone boundaries. In this approximate band model, the Γ valley and the two degenerate valence bands are also taken as parabolic, with the effective masses $m_0 = 0.1m$ and $m_h = 0.34m$, respectively.

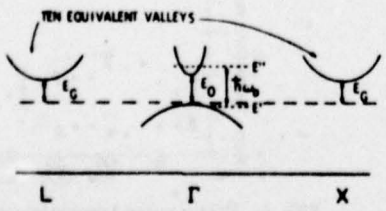


FIG. 6. Reduced Ge energy-band structure.

C. Physics of nonlinear light absorption in Ge

When an excitation pulse is incident on a Ge sample, a portion is reflected. The unreflected portion of the excitation pulse enters the bulk of the crystal where most of it is absorbed. The light entering the bulk of the crystal is absorbed primarily by two processes. In the first process, "direct optical absorption," a quanta of light from the excitation pulse is absorbed by an electron which makes a transition from near the top of the valence band to the conduction-band valley near Γ, leaving behind a hole in the valence band. Such a process is allowed since the energy of the light quanta $\hbar\omega_0$ is greater than the direct band gap E_0. Once there are electrons in the conduction band and holes in the valence band, "free-carrier absorption" is possible. Free-carrier absorption is the process whereby an electron in any of the conduction-band valleys makes a transition to a state higher in that same valley by means of the simultaneous absorption of a quanta of light and the absorption or emission of a phonon (optical or acoustical). An identical process occurs for holes in the valence band. The rate for direct absorption is usually larger than that for free-carrier absorption; however, the rate at which direct absorption events can occur decreases as the number of occupied electron states in the Γ valley increases. On the other hand, the rate for free-carrier absorption events increases as the number of electrons (holes) in the conduction (valence) band increases.

Indirect phonon-assisted interband absorption processes, which involve the transition of an electron from the valence band near Γ to either the L or X conduction-band valleys by the simultaneous

absorption of a photon and the absorption or emission of a phonon, are not important in our problem. The probability that an electron will reach the L or X valley by means of a real optical transition to the Γ valley followed by a phonon-assisted scattering to one of these side valleys is much greater than the probability that the electron will reach the same valley by means of a second-order phonon-assisted transition. For this reason, indirect optical-absorption events may be ignored.[12]

Because of the small number of electron states available in the Γ valley, one might conclude that the direct transitions are saturated at relatively low pulse energies. In fact, this is not the case since an electron in the Γ valley will rapidly emit or absorb a phonon and make a transition to one of the valleys at X or L. A rate calculation in Appendix A shows that the rate for electron transitions from the Γ valley to the L valley is greater than 10^{14} sec^{-1} since $\hbar\omega_0 > E_{L,X} + \max(\hbar\Omega_{\mu q})$ and since the electron-phonon coupling constants in Ge are relatively large. Here $E_{L,X}$ refers to the indirect gaps and $\Omega_{\mu q}$ to the phonon frequency of mode μ and momentum $\vec{q}$. Thus electrons are emptied from the central valley to the side valleys at a rate that is larger than the direct absorption rate, and any decrease in the number of states available for direct absorption is ultimately determined by the buildup of the electron populations in the X and L valleys.

Owing to the narrow bandwidth of the excitation pulse, only narrow regions of states in the central valley of the conduction and valence bands are optically coupled by direct transitions. Thus nonequilibrium carriers might be thought to occupy very localized regions within the conduction bands and valence band, respectively. However, because each carrier moves in the screened Coulomb field of the other carriers, carrier-carrier scattering events will occur. These events include electron-electron, electron-hole, and hole-hole collisions. The rate for such events is large. These collisions ensure that the electron and hole distributions will be Fermi-like. They also ensure that the Fermi distribution for holes and the Fermi distribution for the electrons will reach a common temperature which is in general different from the lattice temperature. Rate calculations in Appendix B using two non-Fermi-like distributions indicate that if they are not Fermi-like, their lifetimes are quite short compared to the inverse of the direct optical-absorption rate. For this reason we take the electron and hole carrier distributions to be Fermi-like at all times.

Electrons located high in a conduction-band valley may relax within that valley by phonon emission. Holes, of course, may also relax by emitting phonons. The effect of this relaxation is to reduce the electron temperature and increase the lattice temperature. The increase in lattice temperature due to these phonon-assisted intraband relaxation processes is significant only at very large pulse energies.

It is important to notice that the total number of carriers is unchanged by free-carrier absorption or phonon-assisted relaxation within the valleys. These processes serve only to elevate or reduce, respectively, the distribution temperature. Only direct absorption will increase the number of carriers. Recombination processes serve to reduce the carrier number.

The recombination processes can be divided into two general categories: radiative and nonradiative. In the present work, nonradiative recombination is much more important in reducing the carrier number than radiative recombination. As the carrier density builds up (as a result of direct absorptions) the plasma frequency of these carriers increases. At sufficiently large plasma frequencies, an electron in the Γ valley can recombine with a hole near the top of the valence bands via emission of a plasmon. Normally, electrons near the conduction-band edge (at Γ) can recombine with holes via emission of plasmons only if the plasma frequency ω_p is larger than the direct gap frequency $E_0/\hbar$ [see Eq. (53)]. However, in our case, the plasma resonance is considerably broadened due to the nature of the direct absorption and subsequent scatterings between and within the Γ valley and the L-X valleys. Therefore, plasmon-assisted recombinations can occur at plasma frequencies lower than $E_0/\hbar$. As the electron and hole populations build up, the plasmon-assisted recombination rate becomes comparable to the direct absorption rate, causing the carrier number to saturate at plasma frequencies much less than $E_0/\hbar$. These plasmon-assisted recombinations can be as important in determining the final number of conduction electrons as the saturation in the total number of available electron states in the conduction band due to the buildup of the electron population.

The collective plasma oscillations have a lifetime that is short compared to a picosecond. The energy lost in the decay of the plasma oscillations is rapidly transferred to single electron and hole states and, thus, ultimately increases the temperature of the carrier distribution. Thus the plasmon-assisted recombinations both limit the number of carriers and raise the electron-hole temperature.

Radiative recombination may be of two types: direct or indirect. The recombination of an electron in the Γ valley of the conduction band with a hole in the valence band by means of emission of

a photon is termed direct; the recombination of an electron in the L or X valley with a hole in the valence band by means of simultaneous emission of a photon and emission (or absorption) of a phonon is termed indirect. Both processes occur on time scales larger than nanosecond, and are not important in our problem.

The diffusion[4] of carriers from the interaction region (focused spot size times sample thickness) also reduces the carrier number. Preliminary diffusion (Boltzmann equation) calculations (to be published) indicate that the number of carriers leaving the interaction region on a picosecond time scale is small. In any case we are able to account for our experimental results without invoking diffusion.

In summary, the following rapid processes are invoked in explaining the passage of our 1.06-μm psec pulses through Ge: particle-particle scattering, electron-phonon scattering, direct and free-carrier absorption, and plasmon-assisted interband electron-hole recombinations. In the calculation, radiative recombination and diffusion are ignored as discussed above.

Physically, the single pulse transmission of Ge as a function of incident pulse energy (see Fig. 2) can be accounted for by direct interband transitions followed by heating of the electron distribution as follows: When the excitation pulse enters the crystal, it is absorbed by direct optical transitions, creating a large number of electrons in the central valley of the conduction band. These electrons are quickly scattered to the side valleys by phonons. Particle-particle scattering events ensure that the carrier distributions are Fermi-like and that both electron and hole distributions have the same temperature. The transmission initially rises due to the saturation of the available optically coupled states as a result of direct absorption. Further increase in the transmission for intensities larger than 10^{14} quanta is hindered as the electrons are heated and removed from these states by free-carrier absorption and plasmon-assisted recombination. For these excitation pulse intensities, electrons in the Γ valley fall to the hole pockets in the valence bands via plasmon emissions at a rate comparable to light absorption thus limiting the carrier density.

After the passage of the excitation pulse, the interaction region of the sample contains a large number of carriers with a high electron-hole temperature. Plasmon-assisted recombinations are essentially turned off when the excitation pulse has passed. This is discussed in detail in Sec. II E. As time progresses, the distribution temperature is reduced by phonon-assisted intraband and intravalley relaxation; the probe pulse interrogates this evolution of the distribution, since it is a sensitive measure of whether the optically coupled states are available for absorption or are occupied.

The probe pulse transmission (see Fig. 3) can be understood in the following way After the passage of the excitation pulse, the electrons (holes) are located high (low) in the valleys because of the high distribution temperature, leaving the states that are optically coupled available for absorption. Hence probe transmission is small. Later, as the distribution cools by means of intraband phonon-assisted transitions and carriers fill the states needed for absorption, the probe transmission increases. Finally, as the distribution temperature cools to near the lattice temperature, the electrons (holes) occupy states near the bottom (top) of the valleys, and states needed for absorption are once again available; the probe transmission then decreases. This is summarized in pictorial terms in Fig. 7.

II. THEORY

A. Electron and hole distributions

In the Introduction, we argued that the electron and hole populations in our problem can be described by two Fermi distributions which can be assigned the same temperature because of the rapid Coulomb collisions. Electrons and holes have distinct Fermi energies. Thus

$$f_{cr}(\vec{k}) = \left[1 + \exp\left(\frac{\mathcal{E}_{cr}(\vec{k}) - \epsilon}{\tau}\right)\right]^{-1},$$

$$\mathcal{E}_0 \leq \mathcal{E}_{cr}(\vec{k}) < +\infty, \quad \text{(1a)}$$

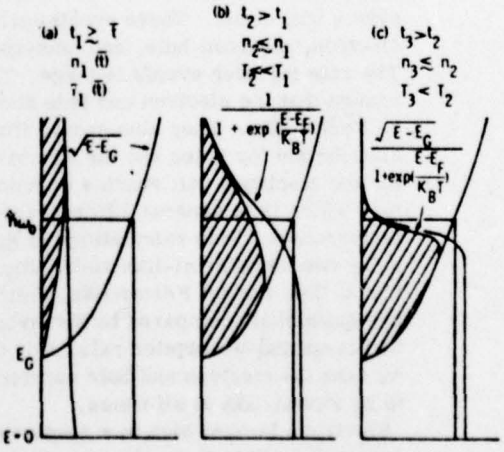


FIG. 7. Schematic diagram for the temporal evolution of the free carriers created by the excitation pulse.

$$f_{cL,X}(\vec{k}) = \left[1 + \exp\left(\frac{\mathcal{E}_{cL,X}(\vec{k}) - \epsilon}{\mathcal{T}}\right)\right]^{-1},$$

$$\mathcal{E}_G \le \mathcal{E}_{cL,X}(\vec{k}) < +\infty, \quad \text{(1b)}$$

and likewise for the holes

$$f_H(\vec{k}) = \left[1 + \exp\left(\frac{\mathcal{E}_H(\vec{k}) + h}{\mathcal{T}}\right)\right]^{-1}, \quad 0 \le \mathcal{E}_H(\vec{k}) < +\infty. \quad \text{(1c)}$$

Here, all energies are normalized with respect to $\hbar\omega_0$, where ω_0 is the circular frequency of the optical pulse. The subscript $c\Gamma$ denotes the Γ valley of the conduction band, cL,X the L-X valleys of the conduction band, and H the holes in the valence bands. ϵ and h are the normalized Fermi energies of the electrons and holes, respectively; and $\mathcal{E}_0$ and $\mathcal{E}_G$ are the normalized energies of the direct and indirect energy gaps, respectively. In the reduced Ge energy-band structure, the indirect energy gap is taken to be $\mathcal{E}_G \equiv \frac{1}{14}(8\mathcal{E}_L + 6\mathcal{E}_X)$, where $\mathcal{E}_L$ is the normalized energy band gap at L and $\mathcal{E}_X$ is the normalized energy band gap at X. As discussed earlier we will treat the L and the X valleys as equivalent and assume that they have spheroid parabolic band structure. This simplifies the algebra considerably, but does not make much difference in the end results. Also note that, due to the rapid phonon-induced scattering of electrons between the central and the side conduction-band valleys (see Appendix A), we have taken the Γ valley and the L- and X-valley Fermi energies to be identical. In the reduced band structure, electron and hole energies are given by

$$\mathcal{E}_{cL,X}(\vec{k}) = (1/\hbar\omega_0)[E_G + \hbar^2(\vec{k} - \vec{k}_{L,X})^2/2m_c]$$

$$\equiv \mathcal{E}_G + \hbar^2(\vec{k} - \vec{k}_{L,X})^2/2\hbar\omega_0 m_c, \quad \text{(2a)}$$

$$\mathcal{E}_{c\Gamma}(\vec{k}) = (1/\hbar\omega_0)(E_0 + \hbar^2k^2/2m_0)$$

$$\equiv \mathcal{E}_0 + \hbar^2k^2/2\hbar\omega_0 m_0, \quad \text{(2b)}$$

$$\mathcal{E}_H(\vec{k}) = \hbar^2k^2/2\hbar\omega_0 m_h, \quad \text{(2c)}$$

and

$$h = E_H/\hbar\omega_0, \quad \epsilon = E_F/\hbar\omega_0, \quad \mathcal{T} = k_B T/\hbar\omega_0. \quad \text{(2d)}$$

Here $\vec{k}$ is a wave vector, k_B Boltzmann's constant, m_0 the effective mass for the Γ valley, m_c the effective mass for both the L and the X valleys, and m_h the effective mass for the heavy-hole bands. $\vec{k}_{L,X}$ refer to the wave vectors for the minima of the conduction bands. The distributions defined above will depend on time through ϵ, h, and $\mathcal{T}$. The Fermi energies h and ϵ are related since the total number of electrons is equal to the total number of holes at all times.

In the present work, it is convenient to measure the electron density in terms of N_0, which is defined by

$$N_0 = \epsilon_\infty m\omega_0^2/4\pi e^2, \quad \text{(3)}$$

where ϵ_∞ is the high-frequency dielectric constant, e the electron charge, and m the bare electron mass. Let us note that due to its smaller effective mass and higher band gap, the Γ valley has a negligible density of electrons compared to the L-X valleys. Thus, neglecting the Γ valley and considering the effective mass of each of the four L valleys to be equal to that of each of the six X valleys, the total electron density is calculated as ten times the density in any one valley. The electron density is given by

$$n = \mathcal{T}^{3/2}\rho_1(z_1), \quad \text{(4a)}$$

where the ρ's and the z's are defined in Table I. Likewise, the hole density is given by

$$n = \mathcal{T}^{3/2}\rho_2(z_2) \quad \text{(4b)}$$

and $\rho_2(z_2) = \rho_1(z_1)$ gives the integral relation between ϵ and h.

By taking the electron and hole distributions to be time dependent only, rather than dependent on

TABLE I. Definitions of integrals.

$$F(x) = (1 + e^x)^{-1}$$

$$\kappa(x) = \int_0^\infty dy\, yF(y + x)$$

$$\rho_i(x) = A_i \int_0^\infty dy\, y^{1/2}F(y + x)$$

$$\sigma_i(x) = A_i \int_0^\infty dy\, y^{1/2}[F(y + x)]^2$$

$$\nu_i(x) = A_i \int_0^\infty dy\, y^{3/2}F(y + x)$$

$$\zeta_j(x) = \int_0^\infty dy \{[(y+1)^{1/2} + (y)^{1/2}]^{4+j} - [(y+1)^{1/2} - (y)^{1/2}]^{4+j}\} \times \left[F\left(x + \left(\frac{y}{\mathcal{T}}\right)\right) - F\left(x + \left(\frac{y+1}{\mathcal{T}}\right)\right)\right]$$

$$A_1 = (5/\pi^2N_0)(2m_c\omega_0/\hbar)^{3/2}$$

$$A_2 = (1/\pi^2N_0)(2m_h\omega_0/\hbar)^{3/2}$$

$$A_3 = (1/2\pi^2N_0)(2m_0\omega_0/\hbar)^{3/2}$$

$$z_1 = (\mathcal{E}_G - \epsilon)/\mathcal{T}$$

$$z_2 = h/\mathcal{T}$$

$$z_3 = (\mathcal{E}_0 - \epsilon)/\mathcal{T}$$

$$z_4 = [\mathcal{E}_0 + (m_h/m_0 + m_h)(1 - \mathcal{E}_0) - \epsilon]/\mathcal{T}$$

$$z_5 = [(m_0/m_0 + m_h)(1 - \mathcal{E}_0) + h]/\mathcal{T}$$

$$z_6 = [\mathcal{E}_0 - (\mathcal{E}_G/2) + (\mathcal{T}_L/2)\ln(A_1/A_2)]/\mathcal{T}_L$$

both time and space, we are ignoring the pulse propagation problem within the interaction region (spot area times the sample thickness) as far as the electronic structure is concerned. This physically reasonable assumption will be justified and its consequences will be detailed in a subsequent paper dealing with the pulse propagation problem. The distributions defined in Eqs. (1a)–(1c) should be viewed as spatial averages of the actual distributions over the interaction region.

We also take the photon density within the interaction region to be time (but not space) dependent: $N_{\vec{q}}^{(\mathrm{rad})} = N_{\vec{q}}^{(\mathrm{rad})}(t)$. In addition, we assume that all quanta have the same wave vector. The radiation energy density at time t is then given by

$$W(t) = \sum_{\vec{q}} \hbar\omega_q N_{\vec{q}}^{(\mathrm{rad})}(t) \simeq \hbar\omega_0 N_0^{(\mathrm{rad})}(t) \equiv \hbar\omega_0 N_0 N(t) ,$$

where $N(t)$ is a normalized "photon" number density. Just as with the electron density, $N(t)$ should be viewed as a spatial average of the actual radiation density over the interaction region.

The following equations (for the electron-hole distributions) summarize the discussion of Sec. I:

$$\frac{d}{dt} f_{c\Gamma}(\vec{k}) = \left(\frac{\partial f_{c\Gamma}}{\partial t}\right)_{DA} + \left(\frac{\partial f_{c\Gamma}}{\partial t}\right)_{FCA} + \sum_j \left(\frac{\partial f_{c\Gamma}}{\partial t}\right)_{\Gamma-j} + \left(\frac{\partial f_{c\Gamma}}{\partial t}\right)_{REL} + \left(\frac{\partial f_{c\Gamma}}{\partial t}\right)_{R} , \tag{5a}$$

$$\frac{d}{dt} f_{cj}(\vec{k}) = -\left(\frac{\partial f_{c\Gamma}}{\partial t}\right)_{\Gamma-j} + \left(\frac{\partial f_{cj}}{\partial t}\right)_{FCA} + \left(\frac{\partial f_{cj}}{\partial t}\right)_{REL} , \tag{5b}$$

$$\frac{d}{dt} f_H(\vec{k}) = \left(\frac{\partial f_H}{\partial t}\right)_{DA} + \left(\frac{\partial f_H}{\partial t}\right)_{FCA} + \left(\frac{\partial f_H}{\partial t}\right)_{REL} + \left(\frac{\partial f_H}{\partial t}\right)_{R} , \tag{5c}$$

where DA represents the direct absorptions; FCA, the free-carrier absorptions; $\Gamma - j$, the phonon-assisted scattering from the Γ valley to the j valley; REL, the intravalley relaxation of electrons and holes by means of emission or absorption of phonons; and R, the direct plasmon-assisted recombination. We note that the direct absorption terms mainly contribute to an increase in the carrier number. The free-carrier absorption contributions serve to increase the distribution temperature. The intravalley phonon-assisted relaxation terms decrease the distribution temperature. The direct recombination terms decrease the electron number and increase the distribution temperature. The $\Gamma - j$ scattering terms do not change the electron number or significantly change the electron temperature; rather, they serve to populate the side valleys in the conduction band.

The behavior of the Fermi distributions for both electrons and holes is determined if ϵ, h, and τ are known at all times. Rather than dealing directly with the distribution functions, we will derive another set of equations from (5a)–(5c), which will describe the behavior of ϵ, h, and τ with time. For this purpose, we consider the time evolution of the electron density, which is given by

$$\frac{dn}{dt} = \sum_{\vec{k}} \frac{d}{dt} f_{c\Gamma}(\vec{k}) + \sum_j \sum_{\vec{k}} \frac{d}{dt} f_{cj}(\vec{k}) , \tag{6a}$$

where j denotes a summation over all conduction-band side valleys. $\sum_{\vec{k}}$ includes the spin summation and the normalization factor, i.e.,[13]

$$\sum_{\vec{k}} = \int \frac{2d\vec{k}}{(2\pi)^3 N_0} .$$

We substitute Eqs. (5a) and (5b) into Eq. (6a), and recognizing that the free-carrier, $\Gamma - j$ scattering, and the relaxation terms do not affect the electron density, we obtain the result

$$\frac{dn}{dt} = \sum_{\vec{k}} \left(\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right)_{DA} + \sum_{\vec{k}} \left(\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right)_{R} . \tag{6b}$$

Inspection of Eq. (4a) shows that n is a function of time via ϵ and τ. Thus we can also write

$$\frac{dn}{dt} = \left(\frac{\partial n}{\partial \epsilon}\right)_{\tau} \frac{d\epsilon}{dt} + \left(\frac{\partial n}{\partial \tau}\right)_{\epsilon} \frac{d\tau}{dt} . \tag{6c}$$

We define

$$\gamma_\epsilon \equiv \left(\frac{\partial n}{\partial \epsilon}\right)_{\tau} \quad \text{and} \quad \gamma_\tau \equiv \left(\frac{\partial n}{\partial \tau}\right)_{\epsilon} . \tag{7}$$

In terms of the integrals defined in Table I, we obtain

$$\gamma_\epsilon(t) = \tau^{1/2}\rho_1(z_1) - \tau^{1/2}\sigma_1(z_1) , \tag{8}$$

and

$$\gamma_\tau(t) = \tfrac{3}{2}\tau^{1/2}\rho_1(z_1) + z_1\tau^{1/2}[\rho_1(z_1) - \sigma_1(z_1)] . \tag{9}$$

Therefore, combining Eqs. (6b) and (6c) we write

$$\gamma_\epsilon(t)\frac{d\epsilon}{dt} + \gamma_\tau(t)\frac{d\tau}{dt} = \sum_{\vec{k}} \frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\bigg|_{DA} + \sum_{\vec{k}} \frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\bigg|_{R} . \tag{10}$$

The fact that the number of electrons equals the number of holes requires that

$$\frac{dn}{dt} = \left(\frac{\partial n}{\partial h}\right)_{\tau} \frac{dh}{dt} + \left(\frac{\partial n}{\partial \tau}\right)_{h} \frac{d\tau}{dt} = \left(\frac{\partial n}{\partial \epsilon}\right)_{\tau} \frac{d\epsilon}{dt} + \left(\frac{\partial n}{\partial \tau}\right)_{\epsilon} \frac{d\tau}{dt} . \tag{11}$$

We define

$$\gamma_{\epsilon h} \equiv -\left(\frac{\partial n}{\partial \epsilon}\right)_{\mathcal{T}} \Big/ \left(\frac{\partial n}{\partial h}\right)_{\mathcal{T}} ;$$

$$\gamma_{\mathcal{T}h} \equiv \left[\left(\frac{\partial n}{\partial \mathcal{T}}\right)_{\epsilon} - \left(\frac{\partial n}{\partial \mathcal{T}}\right)_{h}\right] \Big/ \left(\frac{\partial n}{\partial h}\right)_{\mathcal{T}} . \tag{12}$$

Then, from (11) and (12) we have

$$\frac{dh}{dt} = -\gamma_{\epsilon h}(t)\frac{d\epsilon}{dt} + \gamma_{\mathcal{T}h}(t)\frac{d\mathcal{T}}{dt} . \tag{13}$$

Expressions for $\gamma_{\epsilon h}$ and $\gamma_{\mathcal{T}h}$ are evaluated by differentiating Eqs. (4a) and (4b). In terms of the integrals defined in Table I, these expressions are

$$\gamma_{\epsilon h}(t) = [\rho_1(z_1) - \sigma_1(z_1)] / [\rho_1(z_1) - \sigma_2(z_2)] \tag{14}$$

and

$$\gamma_{\mathcal{T}h}(t) = z_2 - z_1\gamma_{\epsilon h}(t) . \tag{15}$$

To this point, we have obtained two differential equations, (10) and (13), relating three unknowns: ϵ, h, $\mathcal{T}$. We wish to find still a third independent differential equation relating these variables. We accomplish this by considering the energy density of the electrons and holes. The energy density of the electrons and holes, normalized by the photon energy $\hbar\omega_0$ and the number density N_0, is given by

$$u = \frac{U}{\hbar\omega_0 N_0} = \sum_j \sum_{\vec{k}} \left(\frac{E_{cj}(\vec{k})}{\hbar\omega_0}\right) f_{cj}(\vec{k}) + 2\sum_{\vec{k}} \left(\frac{E_H(\vec{k})}{\hbar\omega_0}\right) f_H(\vec{k}) + \sum_{\vec{k}} \left(\frac{E_{c\Gamma}(\vec{k})}{\hbar\omega_0}\right) f_{c\Gamma}(\vec{k}) \tag{16}$$

and du/dt by

$$\frac{du}{dt} = \sum_{\vec{k}} \mathcal{E}_{c\Gamma}(\vec{k})\frac{df_{c\Gamma}(\vec{k})}{dt} + \sum_{\vec{k}}\sum_j \mathcal{E}_{cj}(\vec{k})\frac{df_{cj}(\vec{k})}{dt} + 2\sum_{\vec{k}} \mathcal{E}_H(\vec{k})\frac{df_H(\vec{k})}{dt} . \tag{17}$$

The energy lost to the lattice due to $\Gamma \to j$ phonon-assisted scattering events is negligible. As discussed earlier, plasmon-assisted recombinations do not change u. Radiative recombinations would decrease the total energy, but they are negligible. The free-carrier absorption and phonon-assisted intravalley relaxation contributions of the Γ valley to the energy are negligible compared to those of the L and X valleys. The following equations summarize these statements:

$$\frac{du}{dt} = \left(\frac{\partial u}{dt}\right)_{DA} + \left(\frac{\partial u}{\partial t}\right)_{FCA} + \left(\frac{\partial u}{\partial t}\right)_{REL} , \tag{18a}$$

where

$$\left.\frac{\partial u}{\partial t}\right|_{DA} = \sum_{\vec{k}} \mathcal{E}_{c\Gamma}(\vec{k})\left(\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right)_{DA} + 2\sum_{\vec{k}} \mathcal{E}_H(\vec{k})\left(\frac{\partial f_H(\vec{k})}{\partial t}\right)_{DA} , \tag{18b}$$

$$\left.\frac{\partial u}{\partial t}\right|_{FCA} = \sum_{\vec{k}}\sum_j \mathcal{E}_{cj}(\vec{k})\left(\frac{\partial f_{cj}(\vec{k})}{\partial t}\right)_{FCA} + 2\sum_{\vec{k}} \mathcal{E}_h(\vec{k})\left(\frac{\partial f_H(\vec{k})}{\partial t}\right)_{FCA} , \tag{18c}$$

$$\left.\frac{\partial u}{\partial t}\right|_{REL} = \sum_{\vec{k}}\sum_j \mathcal{E}_{cj}(\vec{k})\left(\frac{\partial f_{cj}(\vec{k})}{\partial t}\right)_{REL} + 2\sum_{\vec{k}} \mathcal{E}_H(\vec{k})\left(\frac{\partial f_H(\vec{k})}{\partial t}\right)_{REL} . \tag{18d}$$

Since $u = u(\epsilon, h, \mathcal{T})$ we can write

$$\frac{du}{dt} = \left(\frac{\partial u}{\partial \epsilon}\right)_{\mathcal{T},h}\frac{d\epsilon}{dt} + \left(\frac{\partial u}{\partial h}\right)_{\mathcal{T},\epsilon}\frac{dh}{dt} + \left(\frac{\partial u}{\partial \mathcal{T}}\right)_{\epsilon,h}\frac{d\mathcal{T}}{dt} . \tag{19a}$$

Making use of the definitions of $\gamma_{\epsilon h}$ and $\gamma_{\mathcal{T}h}$, we obtain

$$\frac{du}{dt} = \left[\left(\frac{\partial u}{\partial \epsilon}\right)_{\mathcal{T},h} - \left(\frac{\partial u}{\partial h}\right)_{\mathcal{T},\epsilon}\gamma_{\epsilon h}\right]\frac{d\epsilon}{dt} + \left[\left(\frac{\partial u}{\partial \mathcal{T}}\right)_{\epsilon,h} + \left(\frac{\partial u}{\partial h}\right)_{\mathcal{T},\epsilon}\gamma_{\mathcal{T}h}\right]\frac{d\mathcal{T}}{dt} . \tag{19b}$$

We define

$$\gamma_{u\epsilon} \equiv \left(\frac{\partial u}{\partial \epsilon}\right)_{\mathcal{T},h} - \gamma_{\epsilon h}\left(\frac{\partial u}{\partial h}\right)_{\mathcal{T},\epsilon} \tag{19c}$$

and

$$\gamma_{u\mathcal{T}} \equiv \left(\frac{\partial u}{\partial \mathcal{T}}\right)_{\epsilon,h} + \gamma_{\mathcal{T}h}\left(\frac{\partial u}{\partial h}\right)_{\mathcal{T},\epsilon} . \tag{19d}$$

Thus combining Eqs. (18a) and (19b), we find that

$$\gamma_{u\epsilon}(t)\frac{d\epsilon}{dt} + \gamma_{u\mathcal{T}}(t)\frac{d\mathcal{T}}{dt} = \left.\frac{\partial u}{\partial t}\right|_{DA} + \left.\frac{\partial u}{\partial t}\right|_{FCA} + \left.\frac{\partial u}{\partial t}\right|_{REL} . \tag{20}$$

Let us note that in the evaluation of $\gamma_{u\epsilon}$ and $\gamma_{u\mathcal{T}}$, the Γ valley can be neglected just as in the case of electron density, and for the same reasons. Thus in terms of the integrals of Table I,

$$\gamma_{u\epsilon}(t) = \tfrac{3}{2}\mathcal{T}^{3/2}\rho_1(z_1)(1+\gamma_{\epsilon h}) + \mathcal{T}^{1/2}\mathcal{E}_G[\rho_1(z_1) - \sigma_1(z_1)] \tag{21}$$

and

$$\gamma_{u\mathcal{T}}(t) = \mathcal{T}^{1/2}\mathcal{E}_G[(\tfrac{3}{2}+z_1)\rho_1(z_1) - z_1\sigma_1(z_1)] + \tfrac{3}{2}\mathcal{T}^{3/2}[\nu_1(z_1) + \nu_2(z_2) + \tfrac{3}{5}z_1\rho_1(z_1)(1+\gamma_{\epsilon h})] . \tag{22}$$

Let us also note that when the Γ valley is neglected,

$$u(t) = \mathcal{T}^{5/2}[\nu_1(z_1) + \nu_2(z_2)] + \mathcal{T}^{3/2}\mathcal{S}_G\rho_1(z_1). \quad (23)$$

Summarizing, Eqs. (10), (13), and (20) are a coupled set of differential equations describing the time evolution of the Fermi energies, ϵ and h and carrier temperature $\mathcal{T}$. For convenience, we summarize our working equations below:

$$\gamma_\epsilon(t)\frac{d\epsilon}{dt} + \gamma_{\mathcal{T}}(t)\frac{d\mathcal{T}}{dt} = \sum_{\vec{k}}\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\bigg|_{DA} + \sum_{\vec{k}}\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\bigg|_{R}, \quad (10)$$

$$\frac{dh}{dt} = -\gamma_{uh}(t)\frac{d\epsilon}{dt} + \gamma_{\mathcal{T}h}(t)\frac{d\mathcal{T}}{dt}, \quad (13)$$

$$\gamma_{u\epsilon}(t)\frac{d\epsilon}{dt} + \gamma_{u\mathcal{T}}(t)\frac{d\mathcal{T}}{dt} = \frac{\partial u}{\partial t}\bigg|_{DA} + \frac{\partial u}{\partial t}\bigg|_{FCA} + \frac{\partial u}{\partial t}\bigg|_{REL}. \quad (20)$$

As the light is absorbed the carrier density and temperature pass through a region in which neither the low-temperature approximation nor the high-temperature approximation holds; as a result we must evaluate these expressions numerically. First, we must turn our attention to calculating the absorption and relaxation rates appearing on the right-hand sides of Eqs. (10) and (20).

B. Direct absorption

In this and Secs. II C and II D we will confine ourselves to presenting only the results in order not to burden the discussion with too much algebra. Direct and free-carrier absorptions in Ge are dealt with extensively in the literature.[14-20] As mentioned earlier, a number of simplifying approximations can be made in our problem. For this reason we discuss the appropriate electron-photon and electron-phonon coupling Hamiltonians in the text, and present a detailed summary of the calculation in Appendix C. The appropriate diagrams describing the perturbation terms used in the calculation are shown in Figs. 8 and 9. We confine ourselves to the lowest-order perturbation terms described by these diagrams, higher-order terms being negligible.

The diagrams in Fig. 8 describe direct absorption and emission (of a quantum $\hbar\omega_0$). The electron-phonon coupling is described by the well-known Hamiltonian

$$H_1 = \sum_{\lambda\vec{q}ss'\vec{k}\vec{k}'}\left(\frac{e}{mc}\right)\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)^{1/2}\langle s'\vec{k}'| \times e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}\,|s\vec{k}\rangle c^\dagger_{s'\vec{k}'}c_{s\vec{k}}b_{\lambda q} + \text{H.c.}, \quad (24)$$

where $|s\vec{k}\rangle$ refers to Bloch states (s is the band index); $\hat{\xi}_\lambda(\vec{q})$ is the unit photon polarization vector; $\vec{p}$ is the electron momentum operator; $c_{s\vec{k}}$ and $b_{\lambda q}$ are the electron and photon annihilation operators, respectively. Electron momenta are generally much larger than the photon momentum; thus

$$\langle s'\vec{k}'|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}\,|s\vec{k}\rangle \simeq \hat{\xi}_\lambda(\vec{q})\cdot\langle s'\vec{k}'|\vec{p}\,|s\vec{k}\rangle. \quad (25a)$$

For direct absorption, generally the absolute square of the interband matrix element of the momentum operator $|\langle c\vec{k}|\vec{p}\,|v\vec{k}\rangle|^2$ is needed. This quantity can be evaluated with the help of the so-called f-sum rule.[21] Near the Γ-valley band edge (BE)

$$|\langle c\vec{k}|\vec{p}\,|v\vec{k}\rangle|^2 \simeq \tfrac{1}{2}m^2E_0(1/m_0 + 1/m_h) \equiv m^2E_0/2m_{BE}. \quad (25b)$$

The laser frequency $\hbar\omega_0$ is such that the direct absorptions will involve the Γ valley and the heavy-hole bands only. The direct absorptions due to the transitions between the light-hole band and the heavy-hole bands are negligible since such transitions are restricted to regions away from the center of the Brillouin zone and thus restricted by the Fermi statistical factors, and since the density of the states of the light-hole band is small.

Using (24) and (25) with the help of the diagrams shown in Fig. 8 and the approximations indicated in Appendix C, one finds

$$\sum_{\vec{k}}\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\bigg|_{DA} = N(t)\alpha_D(t), \quad (26)$$

$$\frac{\partial u}{\partial t}\bigg|_{DA} = N(t)\alpha_D(t), \quad (27)$$

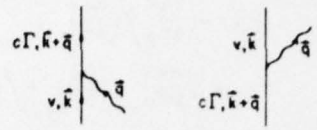


FIG. 8. Direct optical transitions. Solid lines refer to electrons or holes, wavy lines to photons.

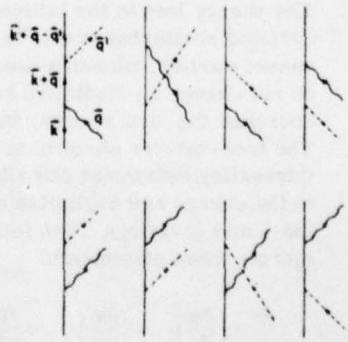


FIG. 9. Free-carrier transitions. Broken lines refer to phonons.

and

$$\left.\frac{\partial N}{\partial t}\right|_{DA} = -N(t)\alpha_D(t), \tag{28}$$

where

$$\alpha_D(t) = \alpha_0[1 - F(z_4) - F(z_5)], \tag{29}$$

and

$$\alpha_0 = \frac{4e^2\mathcal{S}_0}{3\epsilon_\infty \hbar^2}\left(\frac{2m_0 m_h \hbar\omega_0(1-\mathcal{S}_0)}{(m_0+m_h)}\right)^{1/2}. \tag{30}$$

z_4, z_5, and F are defined in Table I. $N(t)$ is the normalized photon density in the interaction region mentioned earlier. $\alpha_D(t)$ is the direct absorption rate. $(\sqrt{\epsilon_\infty}/c)\alpha_0$ is the usual absorption coefficient (per unit length) observed in Beer's-law region[7] (for the light frequency ω_0).

In Fig. 10, α_D/α_0 is plotted as a function of electron-hole temperature for fixed electron-hole densities. The effective masses and the energy-gap parameters are the same for all the curves in the figure.

We observe that as the density decreases, α_D/α_0 approaches 1 for all $\mathcal{T}$:

$$\lim_{n\to 0} \alpha_D = \alpha_0. \tag{31}$$

Also α_D has the same behavior for all densities as $\mathcal{T}\to\infty$, which is to be expected since as $\mathcal{T}$ increases the probability for the occupancy of a given conduction state decreases to essentially zero, thus

$$\lim_{\mathcal{T}\to\infty} \alpha_D = \alpha_0. \tag{32}$$

As $\mathcal{T}\to 0$, α_D shows interesting structure. We see that there are two critical densities such that

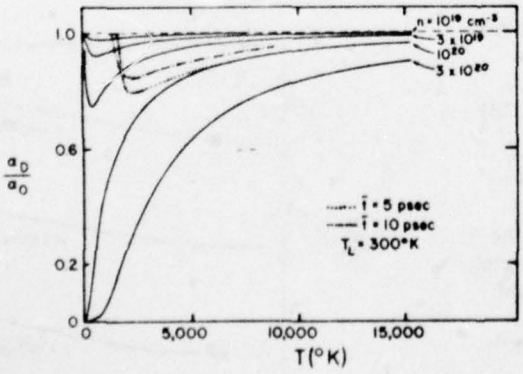


FIG. 10. Direct absorption coefficient vs electron-hole temperature. The curves are generated for the gap parameters corresponding to $T_L = 300°K$. The dotted and dashed curves refer to possible trajectories in this n, $\mathcal{T}$ space as the pulse passes through the sample.

$$\lim_{\mathcal{T}\to 0} \alpha_D = \begin{cases} \alpha_0 & \text{if } n < n_1, \\ \frac{1}{2}\alpha_0 & \text{if } n = n_1, \\ 0 & \text{if } n_1 < n < n_2, \\ -\frac{1}{2}\alpha_0 & \text{if } n = n_2, \\ -\alpha_0 & \text{if } n_2 < n, \end{cases} \tag{33}$$

where n_1 and n_2 are given by

$$n_1 N_0 = \frac{2}{3\pi^2\hbar^3}\left(\frac{2m_0 m_h}{m_0+m_h}(\hbar\omega_0 - E_0)\right)^{3/2} \tag{34a}$$

and

$$n_2 N_0 = \frac{10}{3\pi^2\hbar^3}\left(\frac{2m_c m_h}{m_0+m_h}(\hbar\omega_0 - E_0) + 2m_c(E_0 - E_G)\right)^{3/2}. \tag{34b}$$

The existence of n_1 and n_2 can easily be inferred from Fig. 5. As $\mathcal{T}\to 0$, electrons collect at the bottom of the conduction-band valleys and holes at the top of the valence bands. When $n < n_1$ and $\mathcal{T}\to 0$, $E_F < E''$ and $E_H > E'$; thus direct absorptions can take place unimpeded and $\alpha_D \to \alpha_0$. When $n > n_1$ and $\mathcal{T}\to 0$, $E_H < E'$ and the valence-band states which can absorb $\hbar\omega_0$ quanta are empty. When $n > n_2$ and $\mathcal{T}\to 0$, $E_F > E''$ and $E_H < E'$, and thus the inverse of absorption, i.e., stimulated emission of $\hbar\omega_0$ quanta occurs, and α_D becomes $-\alpha_0$. The two densities n_1 and n_2 are large. For typical values of the effective masses and energy gaps, such as $m_0 \simeq 0.1$, $m_h \simeq 0.34$, $m_c \simeq 0.22$, $E_G \sim 0.7$ eV and $E_0 \sim 0.9$ eV, and $\hbar\omega_0 = 1.17$ eV, we find $n_1 N_0 \sim 4\times 10^{19}$ cm^{-3} and $n_2 N_0 \sim 2\times 10^{21}$ cm^{-3}. In the experiment, the maximum electron densities achieved remain below n_2. Finally, we observe that for $n < n_1$, the minimum of α_D occurs at finite temperatures.

As we shall see later, the direct absorptions remain dominant compared to the free-carrier absorptions for all light intensities used in the experiment, and therefore, the qualitative features of the experimental results can easily be inferred from a diagram such as the one in[22] Fig. 10. $\epsilon_\infty \simeq 16$ for Ge.[23] Thus, for the effective-mass and energy-gap values quoted above, $\alpha_0 \sim 9\times 10^{13}$ sec^{-1}. The usual (Beer's law) linear absorption coefficient is obtained from α_0 by dividing by the speed of light in the absorbing medium, i.e.,

$$\alpha_{\text{Beer}} = (\sqrt{\epsilon_\infty}/c)\alpha_0 = 1.2\times 10^4 \text{cm}^{-1},$$

in agreement with the well-known value.

C. Free-carrier absorption

The electron-phonon interaction is described by the Hamiltonian

$$H_2 = \sum_{\mu,\vec{q}} \sum_{s,\vec{k},\vec{k}'} \left(\frac{\hbar}{2\rho\Omega_{\mu\vec{q}}}\right)^{1/2} \times Q_\mu \langle s\vec{k}' | e^{i\vec{q}\cdot\vec{r}} | s\vec{k}\rangle c^\dagger_{s\vec{k}'} c_{s\vec{k}} a_{\mu\vec{q}} + \text{H.c.}, \tag{35}$$

where μ varies over the phonon modes, ρ is the Ge mass density, $\Omega_{\mu\vec{q}}$ is the phonon frequency, Q_μ is the electron-phonon coupling coefficient, and $a_{\mu\vec{q}}$ is the phonon annihilation operator. We note that for the Bloch states

$$\langle s\vec{k}' | e^{i\vec{q}\cdot\vec{r}} | s\vec{k}\rangle \simeq \delta^3(\vec{k}' - \vec{k} - \vec{q}'), \tag{36a}$$

$$\langle s\vec{k}' | \vec{p} | s\vec{k}\rangle \simeq \hbar\vec{k}\delta^3(\vec{k}' - \vec{k}) , \tag{36b}$$

where s refers to the conduction or the valence bands. Figure 9 illustrates typical free-carrier absorption diagrams. There are altogether 16 such diagrams. With the help of such diagrams and (35)–(36b), the rates $(\partial N/\partial t)$ and $(\partial u/\partial t)$ due to free-carrier absorption can be calculated as summarized in Appendix C. In those calculations, two plausable assumptions are made: (a) the same electron-phonon coupling coefficients are taken for the L and the X valleys; (b) the same coupling coefficients are used for the electrons and the holes. This procedure is followed for the purpose of simplifying the calculations, and can be justified only *a posteriori*, after comparison with the experimental results. Let ϕ_j be the angle between $\vec{q}'$ and the jth valley direction. The electron-phonon coupling coefficients are then as follows. For the longitudinal optical phonons[24]

$$Q_{LO} = Q_0 \cos\phi_j ; \tag{37a}$$

for transverse optical phonons

$$Q_{TO} = Q_0 \sin\phi_j ; \tag{37b}$$

for longitudinal acoustical phonons

$$Q_{LA}(\vec{q}') = q'(\Lambda_d + \Lambda_u \cos^2\phi_j) ; \tag{37c}$$

for transverse acoustical phonons

$$Q_{TA}(q') = -q'\Lambda_u \cos\phi_j \sin\phi_j . \tag{37d}$$

Q_0 is the electron–optical-phonon coupling constant. This coupling constant is taken to be dependent on lattice temperature: $Q_0 = 6 \times 10^{-4}$ at 300 °K and $Q_0 = 2 \times 10^{-4}$ at 77 °K. Λ_d and Λ_u are the same deformation potential constants as in Herring and Vogt.[25] Also, we assume a single reststrahl frequency for the optical phonons, and make a linear approximation for the acoustical phonons[26]:

$$\Omega_{LO,TO} \simeq \Omega_0, \tag{38a}$$

$$\Omega_{LA,TA} \simeq c_A q' . \tag{38b}$$

Free-carrier absorption depends upon the electron population. The Γ-valley electron population is negligible compared to that of the L and X valleys; therefore, the Γ valley is ignored in the FCA calculation. We also assume in Appendix C that each electron after a phonon-assisted photon absorption remains in the same valley, since the intervalley scattering rates among the L and X valleys are small.[27]

With these approximations and others indicated in Appendix C, we obtain the following expressions for the absorption rates:

$$\left.\frac{\partial N}{\partial t}\right|_{FCA} = -N(t)\alpha_{FCA}(t) , \tag{39}$$

$$\left.\frac{\partial u}{\partial t}\right|_{FCA} = N(t)\alpha_{FCA}(t) , \tag{40}$$

where

$$\alpha_{FCA}(t) = \{\alpha_1[1 + 2(\omega_0/\Omega_0)\mathcal{T}_L] + \alpha_2\mathcal{T}_L\} \times [\zeta_0(z_2) + 5(m_c/m_h)^2\zeta_0(z_1)] + \alpha_3[\zeta_1(z_2) + 5(m_c/m_h)^{5/2}\zeta_1(z_1)] , \tag{41a}$$

$$\alpha_1 = e^2Q_0^2m_h^2/3\pi^2\epsilon_\infty\rho\hbar^4\Omega_0 , \tag{41b}$$

$$\alpha_2 = 2e^2m_h^2\Lambda^2\omega_0/3\pi^2\epsilon_\infty\rho c_A^2\hbar^4 , \tag{41c}$$

$$\alpha_3 = 4e^2m_h^2\Lambda^2(2m_h\hbar\omega_0)^{1/2}/15\pi^2\epsilon_\infty\rho c_A\hbar^5 , \tag{41d}$$

$$\Lambda^2 = \Lambda_d^2 + \tfrac{2}{3}\Lambda_d\Lambda_u + \tfrac{1}{3}\Lambda_u^2 . \tag{41e}$$

For the definition of ζ functions see Table I. Here, $\mathcal{T}_L = k_BT_L/\hbar\omega_0$ is the normalized lattice temperature. For Ge, $\epsilon_\infty \simeq 16$, $\rho \simeq 5.3$ g/cm^3, $\Lambda_d \simeq -3.4$ eV, $\Lambda_u \simeq 17$ eV, $Q_0 \simeq 6 \times 10^{-4}$ erg/cm, $\hbar\Omega_0 \simeq 1.88 \times 10^{-2}$ eV, and $c_A \simeq 5.4 \times 10^5$ cm/sec at 300 °K. Using these values we find $\alpha_1 \approx 10^{11}$ sec^{-1}, $\alpha_2 \approx 2 \times 10^{13}$ sec^{-1}, and $\alpha_3 \approx 10^{11}$ sec^{-1}. Although α_2 seems to be quite large compared to α_1 and α_3, actually α_2 occurs in α_{FCA} as $\alpha_2\mathcal{T}_L$, and since generally $\mathcal{T}_L \sim 10^{-2}$, $\alpha_2\mathcal{T}_L$ is of the same order as α_3. Note that

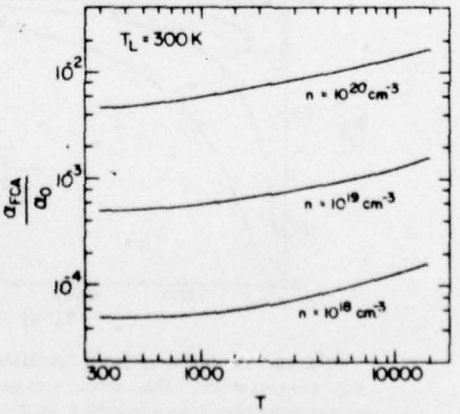


FIG. 11. Free-carrier absorption coefficient vs electron-hole temperature.

α_1, α_2, α_3 are quite small compared to $\alpha_0 \sim 9 \times 10^{13}$ sec^{-1}.

Plots of α_{FCA} as a function of $(\hbar\omega_0/k_B)\mathcal{T}(=T)$ for fixed nN_0 are given in Fig. 11. The behavior of α_{FCA} can easily be deduced from (41) for low- and high-temperature limits. In the low-temperature limit (for fixed n)

$$\zeta_0(z_2) + 5\left(\frac{m_c}{m_h}\right)^2 \zeta_0(z_1) \simeq 8nN_0\left\{\frac{1}{A_2}\left[1+\left(\frac{3n}{2A_2}\right)^{2/3}\right]^{3/2} + \frac{5m_c^2}{m_h^2 A_1}\left[1+\left(\frac{3n}{2A_1}\right)^{2/3}\right]^{3/2}\right\}. \quad (42)$$

For the values of the density that occur in the experiment $3n/2A_1 < 1$ and $3n/2A_2 < 1$, and therefore

$$\zeta_0(z_2) + 5\left(\frac{m_c}{m_h}\right)^2 \zeta_0(z_1) \simeq 8n\left(\frac{1}{A_2} + \frac{5m_c^2}{m_h^2 A_1}\right) \approx 96n \quad (43)$$

and $\alpha_{FCA} \propto n$. At large temperatures, i.e., as $\mathcal{T} \to \infty$,

$$\zeta_0(z_2) + 5\left(\frac{m_c}{m_h}\right)^2 \zeta_0(z_1) - \frac{64}{\sqrt{\pi}}\left(\frac{1}{A_2} + \frac{5m_c^2}{A_1 m_h^2}\right) n\mathcal{T}^{1/2} \approx 500n\mathcal{T}^{1/2} \quad (44a)$$

and

$$\zeta_1(z_2) + 5\left(\frac{m_c}{m_h}\right)^{5/2} \zeta_1(z_1) - 120\left[\frac{1}{A_2} + 5\left(\frac{m_c}{m_h}\right)^{5/2}\right] n\mathcal{T} \approx 1000n\mathcal{T}. \quad (44b)$$

Therefore, as temperature increases, for a given n, α_{FCA} begins to vary linearly with $\mathcal{T}$. We also observe that in both temperature limits $\alpha_{FCA} \propto n$.

D. Phonon-assisted relaxation

We are particularly interested in electron-hole *energy* relaxation via phonon emission and absorption, as it relates to the electron-hole temperature relaxation. The diagrams in Fig. 12 show phonon emission and absorption. As discussed earlier, electron and hole transitions are assumed to be intraband and intravalley in phonon-assisted relaxation. The calculation of $(\partial u/\partial t)_{REL}$ from the diagrams of Fig. 12 is quite similar to the FCA calculation. As shown in Appendix D, we find

FIG. 12. Phonon-assisted relaxation.

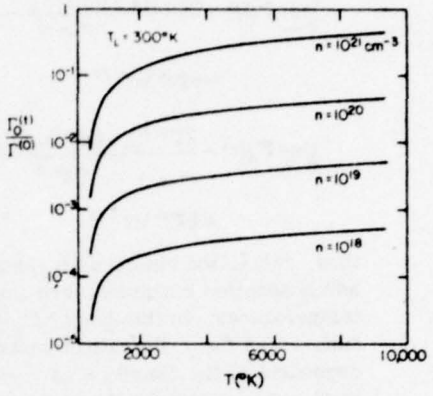


FIG. 13. Optical-phonon-assisted relaxation vs electron-hole temperature.

$$\left(\frac{\partial u}{\partial t}\right)_{REL} = -[\Gamma_0(t) + \Gamma_A(t)], \quad (45)$$

where

$$\Gamma_0(t) = \Gamma^{(0)}\mathcal{T}(\mathcal{T} - \mathcal{T}_L) \times [\ln(1+e^{-z_2}) + 5(m_c/m_h)^3 \ln(1+e^{-z_1})], \quad (46a)$$

$$\Gamma_A(t) = \Gamma^{(a)}\mathcal{T}^2(\mathcal{T} - \mathcal{T}_L)[\kappa(z_2) + 5(m_c/m_h)^4 \kappa(z_1)], \quad (46b)$$

$$\Gamma^{(0)} = 2Q_0^2 m_h^3 \omega_0/\rho\pi^3 N_0 \hbar^4, \quad (46c)$$

$$\Gamma^{(a)} = 16\Lambda^2 m_h^4 \omega_0^2/\rho\pi^3 N_0 \hbar^5, \quad (46d)$$

and $\kappa(x)$ is defined in Table I. $\Gamma_0(t)$ is the relaxation rate due to the optical phonons, $\Gamma_A(t)$, due to the acoustical phonons. For Ge, $\Gamma^{(0)} \approx 2 \times 10^{10}$ sec^{-1} and $\Gamma^{(a)} \approx 5 \times 10^{10}$ sec^{-1} at 300 °K. Owing to the (linear) dispersion of acoustical phonons, the temperature dependence of $\Gamma_0(t)$ and $\Gamma_A(t)$ is different. Plots of $\Gamma_0(t)$ and $\Gamma_A(t)$ vs $(\hbar\omega_0/k_B)\mathcal{T}$ $(= T)$ are given in Figs. 13 and 14. For large temperatures (fixed n)

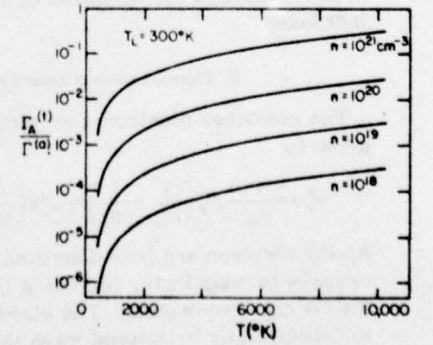


FIG. 14. Acoustical-phonon-assisted relaxation vs electron-hole temperature.

$$\lim_{\mathcal{T}\to\infty}\Gamma_0(t)\to\frac{2\Gamma^{(0)}[1+(m_c/m_h)^{3/2}]\mathcal{T}^{1/2}n}{A_2\sqrt{\pi}}$$

$$\approx 9\Gamma^{(0)}n\mathcal{T}^{1/2}\ , \tag{47a}$$

$$\lim_{\mathcal{T}\to\infty}\Gamma_A(t)\to\frac{2\Gamma^{(a)}[1+(m_c/m_h)^{5/2}]\mathcal{T}^{3/2}n}{A_2\sqrt{\pi}}$$

$$\approx 8\Gamma^{(a)}n\mathcal{T}^{3/2}\ ; \tag{47b}$$

thus, relaxation via acoustic-phonon emission and absorption becomes more important at high temperatures. In this limit, $\Gamma_0+\Gamma_A\propto n$. Since $\lim u\propto n$ as $\mathcal{T}\to\infty$, the rate of relaxation of $\mathcal{T}$ is independent of the density n as $\mathcal{T}\to\infty$. On the other hand, when $\mathcal{T}$ is near $\mathcal{T}_L$ (but n is much greater than the thermal equilibrium density),

$$\Gamma_0(t)\sim(3/2A_2)^{2/3}[1+\tfrac{1}{5}(A_1/A_2)^{4/3}]$$

$$\times\Gamma^{(0)}n^{2/3}(\mathcal{T}-\mathcal{T}_L)\approx 7\Gamma^{(0)}n^{2/3}(\mathcal{T}-\mathcal{T}_L)\ , \tag{48a}$$

$$\Gamma_A(t)\sim\tfrac{1}{2}(3/2A_2)^{4/3}[1+(1/5^{5/3})(A_1/A_2)^{4/3}]$$

$$\times\Gamma^{(a)}n^{4/3}(\mathcal{T}-\mathcal{T}_L)\approx 8\Gamma^{(a)}n^{4/3}(\mathcal{T}-\mathcal{T}_L)\ . \tag{48b}$$

Thus the density dependence of $\Gamma_0(t)$ and $\Gamma_A(t)$ is different. Note also that the relaxation of $\mathcal{T}$ is density dependent since $(\partial u/\partial\mathcal{T})_n\propto n^{1/3}$ in this limit.

The specific heat of the lattice is approximately constant for Ge.[23] Therefore

$$\frac{du_L}{dt}=-\left(\frac{\partial u}{\partial t}\right)_{\text{REL}}=c_L\frac{d\mathcal{T}_L}{dt}\ , \tag{49}$$

where $c_L=C_L/N_0k_B$ and C_L is the specific heat of the Ge lattice in erg/cm^3 °K.[28] Thus

$$\frac{d\mathcal{T}_L}{dt}=\frac{1}{c_L}\left[\Gamma_0(t)+\Gamma_A(t)\right]\ . \tag{50}$$

The energy-gap parameters $\mathcal{E}_G$ and $\mathcal{E}_0$ are functions of $\mathcal{T}_L$. Therefore relaxation, in addition to retarding the growth of $\mathcal{T}$, will also influence n, ϵ, etc., through the variation of $\mathcal{E}_G$ and $\mathcal{E}_0$ as $\mathcal{T}_L$ increases.

E. Plasmon-assisted recombination

The combined electron-hole plasma frequency is given by

$$\omega_p^2=\frac{4\pi e^2n}{\epsilon_\infty}N_0\left(\frac{1}{m_c}+\frac{1}{m_h}\right)=\omega_0^2n\left(\frac{m}{m_c}+\frac{m}{m_h}\right)\ . \tag{51}$$

As the electron and hole densities build up ω_p increases in magnitude, becoming comparable to the Ge gap frequencies. The plasmon resonance is substantially broadened when the excitation pulse is passing through the sample. This is due to the fact that direct absorption populates only the Γ valley. Therefore, the Fermi energy of the Γ-valley electrons is perturbed relative to the Fermi energy of the L-X valley electrons when the excitation pulse is on. This relative perturbation is rapidly damped, as the two Fermi energies try to rapidly equalize by means of the phonon-assisted intervalley scatterings between the Γ valley and the side valleys. It is just such rapid damping that causes the substantial broadening of the plasmon resonance. Note that Coulomb collisions (considered in Appendix B) are ineffective in transferring electrons between the Γ valley and the side valleys, and thus would not damp the relative perturbation of the Fermi energies referred to above. Electrons and holes can of course relax via intravalley and intraband phonon emissions, and thus perturb the Fermi energies. But these are slow processes compared to the Γ-L,X scattering. Their contribution to the plasmon resonance broadening is negligible.

As a result of the broadening of the plasmon resonance, plasmons have short lifetimes and plasmon emission will essentially be spontaneous rather than stimulated. An electron in the Γ valley will recombine with a hole near the top of the valence bands via spontaneous emission of a plasmon. Largely long-wavelength plasmons will be emitted.[29]

For plasmons which have long wavelengths and *infinite* lifetime, the electron-plasmon coupling (in "random-phase approximation") is given by the Hamiltonian[30]

$$H_3\propto -i\sum_{s\vec{k},s'\vec{k}',\vec{q}}\left(\frac{2\pi e^2\omega p}{q^2\epsilon_\infty}\right)^{1/2}$$

$$\times\langle s\vec{k}|e^{i\vec{q}\cdot\vec{r}}|s'\vec{k}'\rangle A_{\vec{q}}c^\dagger_{s\vec{k}}c_{s'\vec{k}'}+\text{H.c.}\ , \tag{52a}$$

where $A_{\vec{q}}$ is the annihilation operator for a plasmon of momentum $\vec{q}$. We are particularly interested in the interband transition terms (i.e., recombination terms) in (52a). Near the Γ-valley band edge, and again for long-wavelength plasmons ($\vec{q}\to 0$),

$$\langle c\vec{k}|e^{i\vec{q}\cdot\vec{r}}|v\vec{k}\rangle\simeq-[(\vec{q}\cdot\vec{P}_{cv})/mE_0]\delta^3(\vec{k}'-\vec{k}+\vec{q}), \tag{52b}$$

where $\vec{P}_{cv}=\langle c\vec{k}|\vec{p}|v\vec{k}\rangle$, the intraband matrix element of the momentum operator as in (25b). From H_3, we find that the recombination rate for a Γ-valley electron with momentum $\vec{k}$ via spontaneous emission of a plasmon (of infinite lifetime) is given by

$$\Gamma_R(\vec{k})=\sum_{v=1}^{2}\sum_{\vec{q}}\left(\frac{4\pi^2e^2\omega_p}{q^2\epsilon_\infty}\right)\left|\frac{\vec{q}\cdot\vec{P}_{cv}}{mE_0}\right|^2$$

$$\times\delta(E_c(\vec{k})-E_v(\vec{k}-\vec{q})-\hbar\omega_p). \tag{53}$$

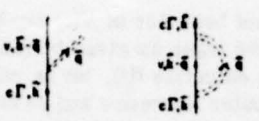


FIG. 15. Plasmon-assisted recombination. Double broken lines refer to plasmons.

However, in our problem, plasmons have a short lifetime, and therefore, the δ function in (53) is replaced by a Lorentzian. More precisely, (53) is replaced by[21]

$$\Gamma_R(\vec{k}) = \left(\frac{2}{\hbar}\right)\sum_{\text{pol}}^{2}\int \frac{d^3\vec{q}}{(2\pi)^3}\left(\frac{4\pi e^2}{q^2\epsilon_\infty}\right)\left|\frac{\vec{q}\cdot\vec{p}_{cv}}{mE_0}\right|^2 \cdot \operatorname{Im}\left(\frac{-1}{\epsilon(\vec{q},[E_{c0}(\vec{k})-E_v(\vec{k}-\vec{q})]/\hbar)}\right), \tag{54}$$

where $\epsilon(\vec{q},\omega)=\epsilon_1(\vec{q},\omega)+i\epsilon_2(\vec{q},\omega)$ is the dielectric function. Equation (54) is obtained from the self-energy diagram shown in Fig. 15. We will approximate Γ_R by the rate for those electrons and holes which are near the band edge at Γ; i.e., we set $k=0$ and $E_{c0}-E_v=E_0+\hbar^2q^2/2m_h$. We will also assume a high-frequency Drude form for the dielectric function:

$$\epsilon(\omega)=1-\frac{\omega_p^2}{\omega(\omega+i/\tau_0)}. \tag{55}$$

$\delta_{pl}=\hbar/2\tau_0$ represents the plasmon resonance broadening mentioned earlier [see Eqs. (56c) and (56d)]. We set $(2/\hbar)\delta_{pl}$ equal to the $\Gamma-L,X$ scattering rate calculated in Appendix A:

$$\frac{1}{\tau_0}=\frac{5(2m_c)^{3/2}(Q_0^2+Q_A^2)(\hbar\omega_0-E_Q)^{1/2}}{2\pi\rho\Omega_0\hbar^3}. \tag{A3}$$

When the plasmon-assisted recombinations are appreciable, electron and hole densities are sufficiently large that $\omega_p\tau_0>1$, and near the plasmon pole ($\omega\sim\omega_p$)

$$\epsilon_1(\omega)\simeq 1-\omega_p^2/\omega^2\simeq(2/\omega_p)(\omega-\omega_p), \tag{56a}$$

$$\epsilon_2(\omega)\simeq 1/\omega_p\tau_0, \tag{56b}$$

$$\operatorname{Im}\left(\frac{-1}{\epsilon(E_0/\hbar+\hbar q^2/2m_h)}\right) \simeq\frac{\hbar\omega_p}{2}\left(\frac{\hbar/2\tau_0}{(\hbar\omega_p-E_0-\hbar^2q^2/2m_h)^2+(\hbar/2\tau_0)^2}\right). \tag{56c}$$

From these approximations and (25b), we obtain

$$\Gamma_R=\frac{2e^2\omega_p}{3\epsilon_\infty E_0 m_{cv}\pi}\int dq\,q^2\,\frac{\hbar/2\tau_0}{(\hbar\omega_p-E_0-\hbar^2q^2/2m_h)^2+(\hbar/2\tau_0)^2}. \tag{56d}$$

Integration over q gives

$$\Gamma_R=\frac{2e^2\omega_p m_h^{3/2}}{3m_{cv}\epsilon_\infty E_0\tau_0\{[(E_0-\hbar\omega_p)^2+(\hbar/2\tau_0)^2]^{1/2}+E_0-\hbar\omega_p\}^{1/2}}. \tag{57}$$

Finally, we can write

$$\left.\frac{dn}{dt}\right|_R=\sum_{\vec{k}}\left.\frac{\partial f_{c\Gamma}}{\partial t}\right|_R=-(n_\Gamma-\bar{n}_\Gamma)\Gamma_R, \tag{58}$$

where

$$n_\Gamma=\Upsilon^{3/2}\rho_3(z_3), \tag{59a}$$

$$\bar{n}_\Gamma=\Upsilon_L^{3/2}\rho_3(z_0) \tag{59b}$$

(see Table I).

F. Rate equations for absorption and transmission

We are now ready to derive the rate equations for describing the absorption and transmission of the excitation pulse. Let us consider the rate equation for $N(t)$ first. It can be written

$$\frac{dN}{dt}=-\alpha(t)N(t)+s(t)-l(t), \tag{60a}$$

$$\alpha(t)=\alpha_D(t)+\alpha_{\Gamma C A}(t), \tag{60b}$$

$s(t)$ is the normalized photon density per unit volume and per unit time that enters the interaction region (source)[32] and $l(t)$ is the normalized photon density per unit volume and time that leaves the interaction region.[33] We must calculate $l(t)$ in order to complete the equation above; however, we would first like to make a few remarks about (60a).

Let the excitation pulse enter the interaction region at $t=0$, and let $N(t=0)=0$. Then, the formal solution of (60a) is

$$N(t)=\int_0^t dt'\,[s(t')-l(t')]\exp\left(-\int_{t'}^{t}\alpha(t'')\,dt''\right). \tag{61}$$

Let us define

$$S(t) = \int_0^t dt'\, s(t'), \tag{62a}$$

$$L(t) = \int_0^t dt'\, l(t'). \tag{62b}$$

$S(t)$ represents the integrated photon density entering the interaction region up to the time t; $L(t)$ represents the integrated photon density that has left the region up to the time $t[S(0)=0, L(0)=0]$. Thus

$$N(t) = S(t) - L(t) - \left[\int_0^t dt'\, \alpha(t')\,[S(t') - L(t')] \times \exp\left(-\int_{t'}^t \alpha(t'')\,dt''\right)\right]. \tag{62c}$$

From this form it is clear that the last term on the right-hand side represents the radiation which is absorbed up to time t:

$$A(t) = \int_0^t dt'\, \alpha(t')[S(t') - L(t')] \exp\left(-\int_{t'}^t \alpha(t'')\,dt''\right). \tag{62d}$$

Thus

$$S(t) = A(t) + N(t) + L(t), \tag{62e}$$

which is an obvious result based on the conservation of energy. In the experiments $L \ll S$, and $L \ll A + N \sim A$. Note that $dA/dt = \alpha(t)N(t)$.

The absorption coefficient α is generally within the range of 10^{14}–10^{13} sec^{-1}. Initially we have $\alpha(t\sim 0) \sim 10^{14}$ sec^{-1}. Clearly, the exponential factor in (61) dominates rapidly [i.e., $\alpha(t)N(t) \gg \dot{N}(t)$], and as $N(t)$ increases from its initial zero value, it will quickly reach the asymptotic value implied by (60a):

$$N(t) = [s(t) - l(t)]/\alpha(t) \tag{63}$$

in a time of order $(1/\alpha) \sim 10^{-14}$ sec. Since the pulse width is much larger than this value, the initial rapid transient behavior of $N(t)$ can be neglected, and $N(t)$ can be taken as given by (63).

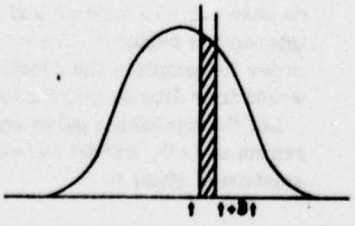


FIG. 16. Partitioning of optical pulse to calculate transmission.

In order to calculate $l(t)$, let us consider a thin slice of the pulse between t and $t+\delta t$ (see Fig. 16) that has entered the crystal. The number of photons in that slice is given by $s\delta t$. The passage time of the slice is $t_p = (\epsilon_\infty)^{1/2}\mathcal{L}/c$, where $\mathcal{L}$ is the width of the interaction region (i.e., the crystal). $t_p \simeq 7\times 10^{-14}$ sec for $\mathcal{L} = 5.5$ μm; thus t_p is much smaller than the pulse width $\bar{t}$. During the passage of the "slice of light" through the crystal, $\alpha(t)$ will not change much since $t_p \ll \bar{t}$, and $\alpha(t)$ can be considered to a good approximation as a function of $A(t)$, and

$$\frac{A(t+t_p) - A(t)}{A(t)} \simeq \frac{t_p}{A(t)}\,\frac{dA}{dt}. \tag{64a}$$

Since $dA/dt = \alpha N$, and from (63) (where we have noted that $l \ll s$) and from (62e) $A \sim S \sim st$

$$\frac{A(t+t_p) - A(t)}{A(t)} \simeq \frac{t_p s(t)}{A(t)} \sim \frac{t_p}{t} \ll 1 \tag{64b}$$

for $t_p \ll t \sim \bar{t}$. The quantity $t_p\alpha > 1$ and multiple reflections can be neglected. For this thin slice, the usual absorption equation

$$\frac{d}{dx}(\delta S) = -\left(\frac{\sqrt{\epsilon_\infty}}{c}\,\alpha\right)\delta S \tag{64c}$$

can be integrated from $x=0$ to $x=\mathcal{L}$ and then multiplied by $(1-r)$ to account for the reflection from the back surface to obtain the density of transmitted quanta,

$$\delta L = (1-r)\delta S(x=\mathcal{L}) = (1-r)s\delta t e^{-t_p\alpha(t)}, \tag{64d}$$

where r is the reflection coefficient. Thus

$$l(t) = (1-r)s(t)e^{-t_p\alpha(t)} \tag{65}$$

and

$$L(t) = \int_0^t dt'\,(1-r)s(t')e^{-t_p\alpha(t')}. \tag{66}$$

The transmission coefficient for the excitation pulse is given by

$$T_R^{(\alpha=0)} = \frac{L(\bar{t})}{S_L} = \frac{1}{S_L}\int_0^{\bar{t}} dt'\,(1-r)s(t')e^{-t_p\alpha(t')}, \tag{67}$$

where S_L represents the normalized density of photons in the laser pulse ($\hbar\omega_0 N_0 S_L$ is the ratio of the laser pulse energy to the interaction volume.) Note that for a rectangular pulse of width $\bar{t}$

$$s(t) = \begin{cases} (1-r)S_L/\bar{t} & \text{for } 0 \le t \le \bar{t}, \\ 0 & \text{for } \bar{t} < t, \end{cases} \tag{68}$$

where $(1-r)$ in front of S_L accounts for the reflection from the front surface of the sample, and thus

$$T_R^{(\mathrm{exc})} = \frac{1}{l}\int_0^{\bar{l}} dl'\,(1-\gamma)^2 e^{-l'\rho_0(t')}. \tag{69}$$

The reflection coefficient was experimentally measured for all intensities of the laser beam used in the experiment, and was found to be nearly constant.[34] It is given by

$$\gamma = (\sqrt{\epsilon_\infty}-1)^2/(\sqrt{\epsilon_\infty}+1)^2 = 0.36. \tag{70}$$

In order to collect together our working equations, we substitute the results given by equations (26), (58), and (63) in (10) and obtain

$$\gamma_\epsilon(t)\frac{d\epsilon}{dt} + \gamma_{\mathcal{T}}(t)\frac{d\mathcal{T}}{dt} = \frac{\alpha_D}{\alpha}(s-l) - (n_\Gamma - \bar{n}_\Gamma)\Gamma_R. \tag{71a}$$

Likewise we substitute (27), (40), (45), and (63) in (20) in order to obtain

$$\gamma_{nn}(t)\frac{d\epsilon}{dt} + \gamma_{n\mathcal{T}}(t)\frac{d\mathcal{T}}{dt}$$

$$= \frac{\alpha_D}{\alpha}(s-l) + \frac{\alpha_{IFA}}{\alpha}(s-l) - [\Gamma_0(t)+\Gamma_A(t)]. \tag{71b}$$

Solving (71a) and (71b) for $d\epsilon/dt$ and $d\mathcal{T}/dt$, taking dh/dt from (13), $d\mathcal{T}_L/dt$ from (50) and $l(t)$ from (65) we have our complete set of working equations:

$$\frac{d\epsilon}{dt} = \left(\frac{\gamma_{n\mathcal{T}}}{\Delta}\right)\left(\frac{\alpha_D}{\alpha}\right)(s-l) - \left(\frac{\gamma_{n\mathcal{T}}}{\Delta}\right)(n_\Gamma - \bar{n}_\Gamma)\Gamma_R$$

$$-\left(\frac{\gamma_{\mathcal{T}}}{\Delta}\right)(s - l - \Gamma_0 - \Gamma_A), \tag{72a}$$

$$\frac{d\mathcal{T}}{dt} = -\left(\frac{\gamma_{nn}}{\Delta}\right)\left(\frac{\alpha_D}{\alpha}\right)(s-l) + \left(\frac{\gamma_{nn}}{\Delta}\right)(n_\Gamma - \bar{n}_\Gamma)\Gamma_R$$

$$+\left(\frac{\gamma_\epsilon}{\Delta}\right)(s - l - \Gamma_0 - \Gamma_A), \tag{72b}$$

$$\frac{dh}{dt} = -\gamma_{\epsilon h}\frac{d\epsilon}{dt} + \gamma_{\mathcal{T}h}\frac{d\mathcal{T}}{dt}, \tag{72c}$$

$$\frac{d\mathcal{T}_L}{dt} = \frac{1}{c_L}(\Gamma_0 + \Gamma_A), \tag{72d}$$

$$l(t) = (1-\gamma)s(t)e^{-l\rho_0(t)}, \tag{72e}$$

where

$$\Delta(t) = \gamma_{n\mathcal{T}}(t)\gamma_\epsilon(t) - \gamma_{nn}(t)\gamma_{\mathcal{T}}(t)$$

$$= \{\mathcal{T}^2/4[\rho_1(z_1) - \sigma_2(z_2)]\}$$

$$\times\{10[\nu_1(z_1)+\nu_2(z_2)][\rho_1(z_1)-\sigma_1(z_1)][\rho_1(z_1)-\sigma_2(z_2)]$$

$$- 9[\rho_1(z_1)]^2[2\rho_1(z_1) - \sigma_1(z_1) - \sigma_2(z_2)]\}. \tag{72f}$$

Given $s(t)$ and $\epsilon(0)$, $h(0)$, $\mathcal{T}(0)$, $\mathcal{T}_L(0)$, these equations can be solved numerically for $\epsilon(t)$, $h(t)$, $\mathcal{T}(t)$, $\mathcal{T}_L(t)$, and $l(t)$ (or L).

G. Integration of dynamical equations

We now make a few qualitative remarks about (72a)–(72e) and present their numerical solutions. At the initial phase of the excitation pulse absorption, $\mathcal{T}$ increases quite rapidly compared to n since after a direct optical transition an electron in the Γ valley scatters to one of the L-X valleys. After such a scattering event an energy of approximately $1-\delta_C$ goes into electronic thermal agitation. Therefore, $\mathcal{T}$ approaches[35] $\frac{1}{3}(1-\delta_C)$ (~1500 °K) rapidly and ϵ decreases to a negative value. In this initial phase, plasmon recombination, phonon relaxation and free-carrier absorption terms are completely neglibible; thus, we may apply a nondegenerate (high temperature) approximation to the integrals in the γ coefficients (see Table II). After this initial state, $\mathcal{T}$ may decrease or increase depending upon the comparative rates of the phonon relaxation, plasmon recombination, and the free-carrier absorption.

As n increases, initially the transmission also increases. Eventually if n is sufficiently large, plasma recombinations begin to compete with direct absorptions and n begins to saturate. The saturation region is reached when

$$(\alpha_D/\alpha)s \sim (n_\Gamma - \bar{n}_\Gamma)\Gamma_R. \tag{73a}$$

This corresponds to $dn/dt \sim 0$. In this region, those electrons which have just made direct optical transitions can recombine with holes via emission of plasmons and become available again for direct absorption. Plasmons rapidly decay and transfer their energy to electrons and holes. Thus, in the saturation region, the main result of direct absorption is the heating of electrons and holes.

TABLE II. High-temperature approximation (low densities).

$F(x) \to e^{-x}$
$\rho_i(z) \to \frac{1}{2}\sqrt{\pi}\,A_i e^{-z}$
$\sigma_i(z) \to (\sqrt{\pi}/4\sqrt{2})A_i e^{-2z}$
$\nu_i(z) \to \frac{3}{4}\sqrt{\pi}\,A_i e^{-z}$
$n \to \frac{1}{2}\sqrt{\pi}\,A_1 \mathcal{T}^{3/2} e^{-z_1} = (\sqrt{\pi}/2)A_2 \mathcal{T}^{3/2} e^{-z_2}$
$\gamma_\epsilon \to n/\mathcal{T}$
$\gamma_{\mathcal{T}} \to (n/\mathcal{T})(1.5 - z_1) \approx -(nz_1/\mathcal{T})$
$\gamma_{nh} \to 1$
$\gamma_{\mathcal{T}h} \to z_2 - z_1$
$\gamma_{nn} \to (n/\mathcal{T})(\delta_G + 3\mathcal{T})$
$\gamma_{n\mathcal{T}} \to (n/\mathcal{T})\{(1.5+z_1)\delta_G + 3\mathcal{T}(2.5+z_1)\} \approx (nz_1/\mathcal{T})(\delta_G + 3\mathcal{T})$
$\Delta \to 3n^2/\mathcal{T}$

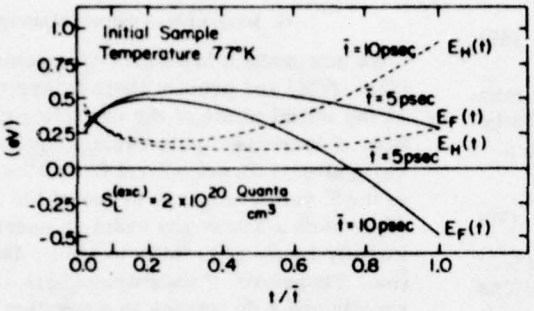


FIG. 17. Temporal behavior of Fermi energies during passage of excitation pulse.

In this region, $\mathcal{T}$ then varies according to

$$\frac{d\mathcal{T}}{dt} \sim \frac{\gamma_s}{\Delta}(s - \Gamma_0 - \Gamma_A) \tag{73b}$$

and increases rather rapidly. In the saturation region, the trajectories of the system remain close to the constant n curves in Fig. 10.

The following several figures present the results of numerically integrating (72a)–(72e) during the period the excitation pulse is present in the sample. All numerical solutions have been generated assuming a rectangular optical pulse of width $\bar{\tau}$. Figures 17 and 18 illustrate the temporal evolution of the Fermi energies of the electrons and holes and of the carrier temperature as an excitation pulse of constant energy traverses the sample. We have presented these curves for several different pulse widths to emphasize that the pulse width has a significant influence on the temporal evolution of these quantities. Note that for the large excitation pulse energy used in these figures the temperature begins near 1500 °K and

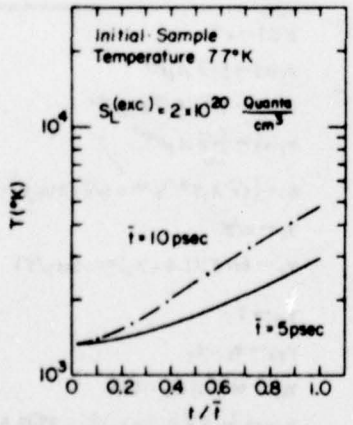


FIG. 18. Temporal behavior of electron-hole temperature.

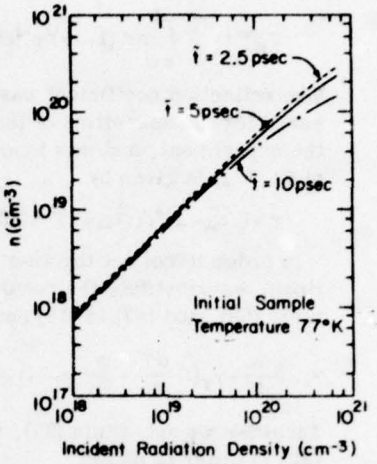


FIG. 19. Electron density vs excitation pulse density for different pulse widths and initial lattice temperature of 77 °K.

rises due to plasmon recombinations as discussed above.

Figures 19–24 show the final carrier number, carrier temperature, and lattice temperature immediately after the passage of the excitation pulse as a function of excitation pulse energy for lattice temperatures of 77 and 297 °K. Again the dependence of these curves on pulse width is emphasized. Notice that, due to plasmon emission, the electron number begins to saturate somewhat and the carrier temperature is elevated for large excitation pulse energies as discussed earlier. Also note that this model does not provide for sufficient lattice heating to account for the surface damage

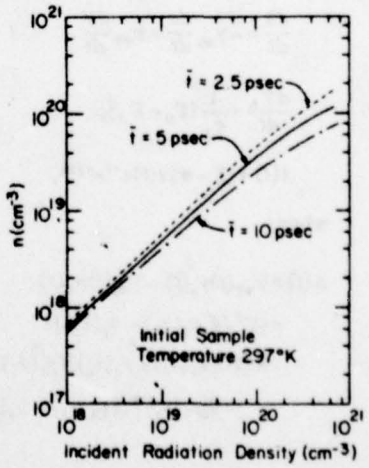


FIG. 20. Electron density vs excitation pulse density for different pulse widths and initial lattice temperature of 297 °K.

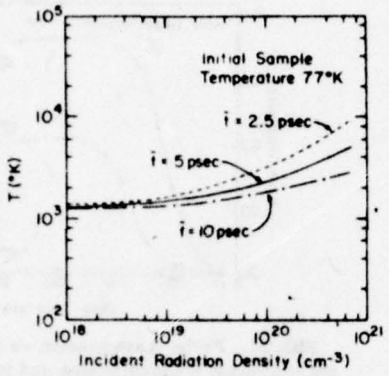


FIG. 21. Electron-hole temperature vs excitation pulse density for different pulse widths and initial lattice temperature of 77 °K.

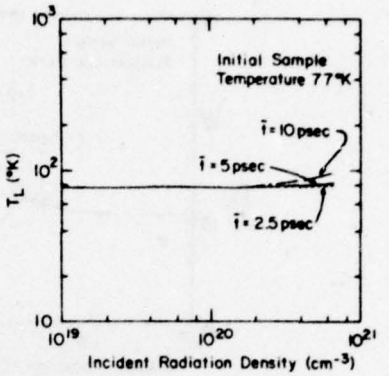


FIG. 23. Lattice temperature vs excitation pulse density for different pulse widths and initial lattice temperature of 77 °K.

we observe at the largest pulse energies. In Figs. 25 and 26, we show the corresponding single pulse, or excitation pulse, transmission. These curves are presented as a function of pulse width for lattice temperatures of 77 and 300 °K.

For $t > \bar{l}$, we set $s = 0, l = 0$. The system begins to evolve according to

$$\frac{d\epsilon}{dt} = \left(\frac{\gamma_I}{\Delta}\right)(\Gamma_0 + \Gamma_A), \tag{74a}$$

$$\frac{d\bar{l}}{dt} = -\left(\frac{\gamma_s}{\Delta}\right)(\Gamma_0 + \Gamma_A). \tag{74b}$$

dT_L/dt and dh/dt are still given by (72c) and (72d). In (74a) and (74b) we have omitted the plasmon-assisted recombinations since, as discussed previously, when the excitation pulse is turned off the plasmon process is likewise rapidly turned off. Notice that in the experiment the largest probe pulse delay was about 300 psec and that direct and indirect optical recombinations are still negligible for such delay times. Since the experimental probe pulse intensities were small compared to the excitation pulse intensities, the system evolves according to (74a) and (74b) and (72c) and (72d) when the probe pulse passes through the interaction region, rather than (72a) and (72b) with $s = s_{\text{probe}}, l = l_{\text{probe}}$.

For a rectangular pulse of width $\bar{l}$, the probe transmission is given by

$$L_{\text{probe}} = \int dt' \, (1 - r) s_{\text{probe}}(t') e^{-t_p \alpha(t')} \tag{75}$$

and the transmission coefficient by

$$T_R^{\text{probe}} = \frac{(1 - r)^2}{\bar{l}} \int_{t_d - \bar{l}/2}^{t_d + \bar{l}/2} dt' \, e^{-t_p \alpha(t')}, \tag{76}$$

where t_d is the delay time of the probe with respect to the excitation pulse. For $t_d > \bar{l}$, T_R^{probe} increases at first due to electron-hole relaxation. As electrons and holes relax via phonon emissions, they clog the optically coupled states E' and E'' in Fig.

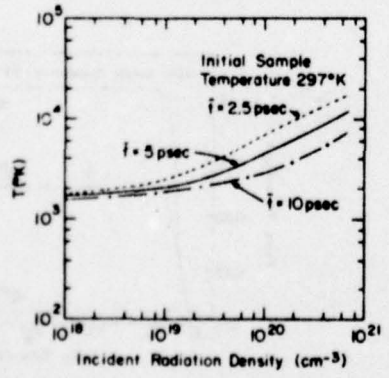


FIG. 22. Electron-hole temperature vs excitation pulse density for different pulse widths and initial lattice temperature of 297 °K.

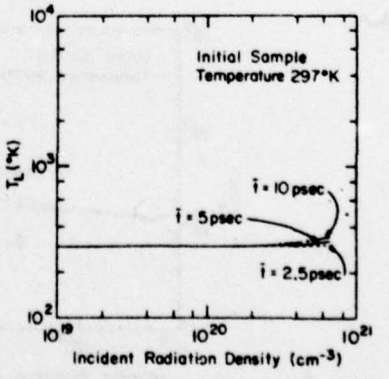


FIG. 24. Lattice temperature vs excitation pulse density for different pulse widths and initial lattice temperature of 297 °K.

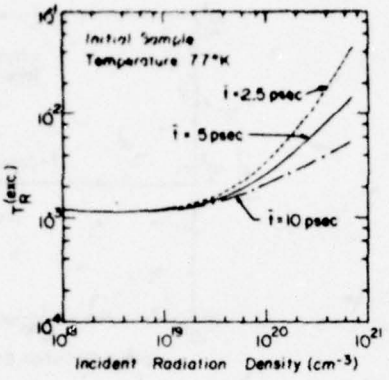


FIG. 25. Excitation pulse transmission vs excitation pulse density for initial lattice temperature of 77 °K.

5 and reduce absorption. Further relaxation can eventually free these states for optical transitions, and probe transmission can decrease. These ideas are summarized in Fig. 7. Figures 27 and 28 show probe transmission versus delay time for different excitation pulse radiation densities and for the two initial sample temperatures 77 and 297 °K. For these solutions, the pulse width is assumed to be 5 psec. As the excitation pulse intensity decreases, the probe transmission curve is generally lowered.[36]

III. COMPARISON WITH EXPERIMENTAL DATA AND CONCLUSIONS

The experimental data and the theoretical curves for the excitation pulse transmission are shown in Fig. 29. The theoretical curves are for a 5-psec pulse width and a 3×10^{-7}-cm^3 interaction volume. This volume corresponds to a focused spot diameter that is in accord with the experimentally estimated value of about 250 μm. The agreement

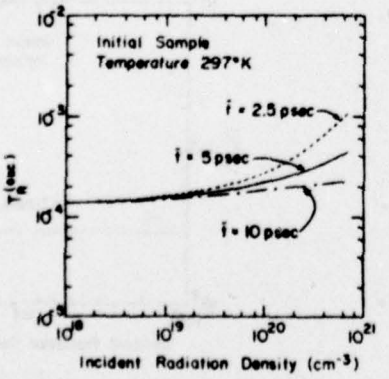


FIG. 26. Excitation pulse transmission vs excitation pulse density for initial temperature of 297 °K.

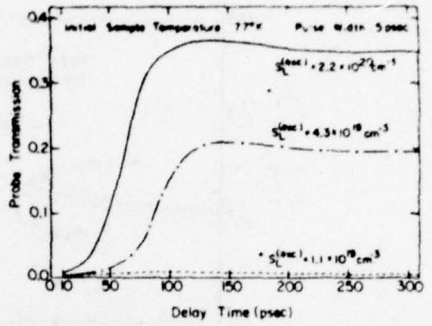


FIG. 27. Probe transmission vs delay time for different excitation pulse densities and initial lattice temperature of 77 °K.

between data and theory is good at lower excitation energies. At the extremely high excitation energies ($\sim 4 \times 10^{14}$ quanta) the theoretical enhanced transmission does not saturate as does the experimental transmission. However, since we are uncertain whether the saturation in experimental transmission is due to bulk effects or due to the onset of surface damage, we shall not dwell on this point.

Figure 30 shows the probe transmission data (minus the zero delay time spike) and the corresponding theoretical curves. The theoretical curves were generated for 2×10^{20}-cm^{-3} incident excitation quanta density, 5-psec pulse width, and 10^{-6}-cm^3 interaction volume. The agreement between the theory and the experimental data is reasonable for both 77 and 297 °K.

The overall good agreement between the theory and the experiment permits us to draw a number of conclusions. An important assumption in our calculations was that the large number of free carriers created by the excitation pulses be described by Fermi-like distributions. Apparently, this assumption works quite well despite the in-

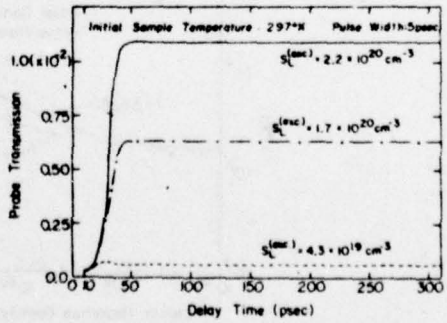


FIG. 28. Probe transmission vs delay time for different excitation pulse densities and initial lattice temperature of 297 °K.

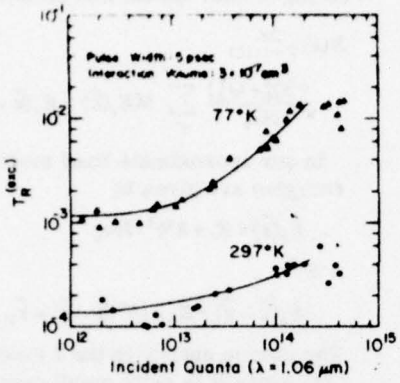


FIG. 29. Theoretical and experimental excitation pulse transmission vs excitation pulse intensity.

tensity and the short duration of the optical pulses. Clearly, in order to observe a large deviation from a Fermi-like distribution in a thin sample, one must go to even shorter pulse widths than those achieved in these experiments. It also appears that if the sample is only a few micrometers thick, then one can average over the spatial effects and obtain good results.

Many-body (plasmon) effects are extremely important in intense nonlinear light absorption owing to the large number of carriers created and the short lifetime of the optically coupled states. Plasmon-assisted recombinations play a role in the behavior of both excitation and probe transmission by saturating the free carrier densities while increasing their temperature. Plasmon-assisted recombinations should have observable consequences in similar experiments with other materials. How they would affect the excitation and probe transmission would depend on the particulars of the energy-band structure of each semiconductor.

Let us emphasize again that the highest plasma

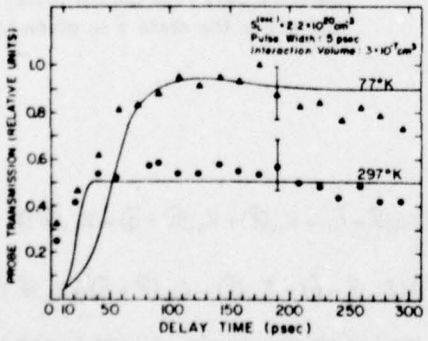


FIG. 30. Theoretical and experimental probe transmission vs delay time.

frequencies achieved in the experiments are actually less than $E_0/2\hbar$. The maximum electron densities occurring in the vicinity of 2×10^{20} cm^{-3}, which corresponds to $(\max\omega_p)/\omega_0 \simeq 0.3$. There is a directly observable quantity, namely the refractive index for the optical frequency ω_0, which is directly related to the ratio ω_p/ω_0. For the electron density quoted above, the maximum change in the refractive index divided by the initial refractive index is given by[37]

$$\max\left(\frac{\Delta_{\text{ref index}}}{\text{ref index}}\right) \simeq -\frac{1}{2}\left(\frac{\max\omega_p}{\omega_0}\right)^2 \simeq -0.05.$$

This is in good agreement with the direct measurements of the refractive index.[4]

In Sec. II, the role of the phonon-assisted relaxation was discussed in detail. We have seen that the reason the relaxation extends over a time span of a hundred picoseconds is due to the fact that the phonon-assisted relaxation in Ge is strongly dependent on the difference between the temperatures of the free carriers and the lattice and due to the existence of conduction band valleys in Ge. In different materials with different energy-band structures and electron-phonon coupling constants, the behavior of the probe pulse transmission would be (and is) drastically different.

As mentioned earlier, the lattice temperature increase is generally small. Here, "small" can mean as much as 10–30 °K at large light intensities. This may be sufficiently large to be experimentally measurable.

An important assumption in the calculation of the phonon-assisted relaxation rates was the fact that phonon distributions were given by the usual Bose distributions. This assumption apparently works well. However, at large intensities (near 10^{15} laser quanta), large numbers of phonons are generated. It is quite possible that a phonon instability may occur, i.e., the number of quanta in some phonon modes may grow to larger values than those given by equilibrium values. This should be investigated in the future.

An interesting feature of the theoretical model as developed is the behavior of the transmission for different pulse widths. One can see from Figs. 25 and 26 that transmission changes significantly near the saturation region when pulse width is increased. Since it is possible to split an optical pulse into separate parts and then rejoin them, it is possible to perform the experiments with variable duration pulses. These and other experiments are currently under way.

ACKNOWLEDGMENTS

We thank F. Hopf for his valuable criticisms, W. P. Latham and R. J. McCarthy for their aid in

computer programming, and J. Bessey, R. Emrick, R. Parmenter, B. O. Seraphin, R. Stark, H. van Driel, and R. Young for helpful comments concerning the manuscript.

APPENDIX A: $\Gamma \rightarrow L,X$ TRANSITIONS VIA PHONON EMISSIONS

Let us consider a particular side valley, say (111) valley. Using the electron-phonon coupling Hamiltonian given by (35), the transition rate for an electron going from the Γ valley to the (111) valley via spontaneous emission of optical and acoustical phonons can be written

$$R(k)=\left(\frac{2\pi}{\hbar}\right)\sum_{\vec{q}',\mu}\left(\frac{\hbar}{2\rho\Omega_{\mu q'}}\right)\times|Q_\mu(q')|^2|\langle c,\vec{k}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|c,\vec{k}\rangle|^2\times\delta(E_c(\vec{k})-E_c(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'}). \quad \text{(A1a)}$$

where $\vec{k}$ is sufficiently small such that $|c,\vec{k}\rangle$ refers to an electron in the Γ valley and $-\vec{q}'$ is sufficiently large and generally pointing in [111] direction such that $|c,\vec{k}-\vec{q}'\rangle$ refers to an electron in the [111] valley. If we let $\vec{k}_{(111)}$ represent the wave vector of the lowest point (in energy) of the [111] valley, then $q' \sim k_{(111)}$, and $q' \gg k$. Since q' is smaller than the reciprocal-lattice vector, $|\langle c,\vec{k}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|c,\vec{k}\rangle|^2 \approx 1$. Electron-phonon coupling parameters $Q_\mu(\vec{q}')$ were discussed in the text. For optical phonons $Q_{LO}=Q_0\cos\phi$ and $Q_{TO}=Q_0\sin\phi$, where ϕ is the angle between $-\vec{q}'$ and $\vec{k}_{(111)}$. Since $-\vec{q}'$ is assumed to be generally in the direction of $\vec{k}_{(111)}$, $Q_{LO}\approx Q_0$ and $Q_{TO}\approx 0$. Thus for optical phonons the right-hand side of (A1a) becomes

$$\frac{\pi Q_0^2}{\rho\Omega_0}\sum_{\vec{q}'}\delta(E_c(\vec{k})-E_c(\vec{k}-\vec{q}')-\hbar\Omega_0). \quad \text{(A1b)}$$

For acoustical phonons, $Q_{LA}=q'(\Lambda_d+\Lambda_u\cos^2\phi)\approx q'(\Lambda_d+\Lambda_u)$ and $Q_{TA}=-q'\Lambda_u\cos\phi\sin\phi\approx 0$. Since $q'\approx k_{(111)}$, $Q_{LA}\approx k_{(111)}(\Lambda_d+\Lambda_u)\equiv Q_A$. Also, $c_A q'\approx c_A k_{(111)}\approx\Omega_0$.[26] Therefore, we obtain the contribution of both optical and acoustical phonons

$$R(k)\underset{\Gamma\rightarrow(111)}{\simeq}\frac{\pi(Q_0^2+Q_A^2)}{\rho\Omega_0}\sum_{\vec{q}'}\delta(E_c(\vec{k})-E_c(\vec{k}-\vec{q}')-\hbar\Omega_0). \quad \text{(A1c)}$$

In our approximate band model, the electron energies are given by

$$E_c(\vec{k})=E_0+\hbar^2k^2/2m_0 \quad \text{(A1d)}$$

and

$$E_c(\vec{k}-\vec{q})=E_G+[\hbar^2(\vec{k}-\vec{q}'-\vec{k}_{(111)})^2]/2m_c. \quad \text{(A1e)}$$

The phonon energy in the δ function can be neglected, since it is quite small compared to electron energies. Letting $-\vec{q}'=\vec{k}_{(111)}+\vec{q}''$, where $q''\ll q', k_{(111)}$, the integral over the δ function is trivial, and yields

$$R(k)\underset{\Gamma\rightarrow(111)}{\simeq}\frac{(2m_c)^{3/2}(Q_0^2+Q_A^2)}{4\pi\rho\Omega_0\hbar^3}\left(E_0+\frac{\hbar^2k^2}{2m_0}-E_G\right)^{1/2}. \quad \text{(A2)}$$

For an electron just raised to the Γ valley by optical transition $E_0+\hbar^2k^2/2m_0\simeq\hbar\omega_0$. Thus, summing over all valleys, we obtain

$$\frac{1}{\tau_0}=\sum_j R_{\Gamma\rightarrow j}\simeq\frac{5(Q_0^2+Q_A^2)(2m_c)^{3/2}}{2\pi\rho\Omega_0\hbar^3}(\hbar\omega_0-E_G)^{1/2}. \quad \text{(A3)}$$

For $m_c=0.22m$, $\hbar\omega_0=1.17$ eV, $E_G\simeq 0.7$ eV, $\rho\simeq 5.3$ g/cm^3, $\hbar\Omega_0\simeq 0.02$ eV, $Q_0\simeq 6\times10^{-4}$ erg/cm, and $Q_A\simeq 2\times10^{-3}$ erg/cm, $1/\tau_0\simeq 1\times10^{14}$ sec$^{-1}\geq$ (the absorption rate).

APPENDIX B: COULOMB THERMALIZATION

In this appendix we will calculate the lifetime of an occupied state, which is finite due to Coulomb scatterings among electrons and holes if the distributions are not Fermi-like. For the sake of simplicity, we will neglect the long-wavelength screening of the Coulomb fields. For an electron state $\vec{k}$ in the ith valley, the scattering rate out of the state $\vec{k}$ is given by

$$\Gamma_{\vec{k}}^{\text{out}}=\left(\frac{2\pi}{\hbar}\right)\int\frac{2d\vec{k}'}{(2\pi)^3}\int\frac{d\vec{q}}{(2\pi)^3}\left(\frac{4\pi e^2}{q^2\epsilon_\infty}\right)^2[1-f_{ci}(\vec{k}-\vec{q})]\times\left(\sum_j f_{cj}(\vec{k}')[1-f_{cj}(\vec{k}'+\vec{q})]\delta(E_{ci}(\vec{k}-\vec{q})-E_{ci}(\vec{k})+E_{cj}(\vec{k}'+\vec{q})-E_{cj}(\vec{k}'))\right.$$
$$\left.+\sum_v[1-f_{hv}(\vec{k}')]f_{hv}(\vec{k}'+\vec{q})\delta(E_{ci}(\vec{k}-\vec{q})-E_{ci}(\vec{k})-E_{hv}(\vec{k}'+\vec{q})+E_{hv}(\vec{k}'))\right). \quad \text{(B1)}$$

j varies over the conduction-band valleys, v over the two valence bands. f_{cj} and f_{hv} are electron and hole distributions, respectively; E_{cj} and E_{hv} are electron and hole energies (E_{hv} is positive).

The scattering rate into the state $\vec{k}$ is given by

$$\Gamma_{\vec{k}}^{\text{in}} = \left(\frac{2\pi}{\hbar}\right) \int \frac{2d^3\vec{k}'}{(2\pi)^3} \int \frac{d^3\vec{q}}{(2\pi)^3} \left(\frac{4\pi e^2}{q^2 \epsilon_\infty}\right) f_{cf}(\vec{k}-\vec{q}) \times \left(\sum_j [1 - f_{cj}(\vec{k}')] f_{cj}(\vec{k}'+\vec{q}) \delta(E_{cf}(k-q) - E_{cf}(k) + E_{cj}(k'+q) - E_{cj}(k')) + \sum_v [1 - f_{hv}(\vec{k}'+\vec{q})] f_{hv}(\vec{k}') \delta(E_{cf}(\vec{k}-\vec{q}) - E_{cf}(\vec{k}) - E_{hv}(\vec{k}'+\vec{q}) + E_{hv}(\vec{k}')) \right). \quad \text{(B2)}$$

Of course, if the distributions are Fermi-like, with a *common* temperature for electrons and holes, then $\Gamma_{\vec{k}}^{\text{in}} = \Gamma_{\vec{k}}^{\text{out}}$, for any state $\vec{k}$. We will evaluate Γ's for two kinds of rather simple distributions which are energy dependent only. To simplify the algebra, we will take the effective masses of electrons and holes to be the same m^* (since $m_c \simeq 0.22m$, $m_h \simeq 0.34m$, this is not too drastic). Also, the total electron density is always taken to be equal to the total hole density.

(a) Let the distributions be Lorentzian with narrow widths, such that

$$f_{cj}(k) = \Delta^2/\{[E_{cj}(\vec{k}) - E_G - \tilde{E}]^2 + \Delta^2\} \simeq \pi\Delta\delta(E_{cj}(\vec{k}) - E_G - \tilde{E}), \quad \text{(B3a)}$$

$$f_{hv}(k) = (5\Delta)^2/\{[E_{hv}(\vec{k}) - \tilde{E}]^2 + (5\Delta)^2\} \simeq 5\pi\Delta\delta(E_{hv}(\vec{k}) - \tilde{E}), \quad \text{(B3b)}$$

i.e., $\delta \equiv \Delta/\tilde{E} \ll 1$.

(i) Let us consider an occupied state $\vec{k}$: $E_{cf}(k) = \tilde{E} + E_G$. Then,

$$\Gamma_{\vec{k}}^{\text{in}} = 0, \quad \text{(B4a)}$$

$$\Gamma_{\vec{k}}^{\text{out}} = (160e^4 m^*/3\hbar^3\epsilon_\infty^2)(1-\delta)^{3/2}/\delta. \quad \text{(B4b)}$$

(ii) For the states $\vec{k}'$ such that $E_{cf}(\vec{k}') < \tilde{E} + E_G - \Delta$ or $E_{cf}(\vec{k}') > \tilde{E} + E_G + \Delta$,

$$\Gamma_{\vec{k}}^{\text{in}} = (80e^4 m^*/3\hbar^3\epsilon_\infty^2)(1-\delta)^{3/2}/\delta, \quad \text{(B5a)}$$

$$\Gamma_{\vec{k}}^{\text{out}} = 0. \quad \text{(B5b)}$$

Setting $m^* \simeq 0.22m$, $\epsilon_\infty = 16$, we find $80e^4 m^*/3\hbar^3\epsilon_\infty^2 \simeq 10^{15}$ sec^{-1}. Thus $\Gamma_{\vec{k}}^{\text{out}}, \Gamma_{\vec{k}'}^{\text{in}} \gg$ (light absorption rate).

(b) Let the distributions be of the following form:

$$f_e(\vec{k}) = \begin{cases} 1 \text{ if } \hbar^2(K - \frac{1}{2}\Delta)^2/2m^* < E_e(\vec{k}) - E_G \\ < \hbar^2(K + \frac{1}{2}\Delta)^2/2m^*, \\ 0 \text{ elsewhere;} \end{cases} \quad \text{(B6a)}$$

$$f_h(\vec{k}) = \begin{cases} 1 \text{ if } \hbar^2(K - \frac{1}{2}\Delta)^2/2m^* < E_h(\vec{k}) \\ < \hbar^2(K + \frac{1}{2}\Delta)^2/2m^*, \\ 0 \text{ elsewhere;} \end{cases} \quad \text{(B6b)}$$

and $2.5\Delta < K < (1.5 + \sqrt{2})\Delta$ (i.e., the width Δ is large).

Let us consider an occupied electron state $\vec{k}$ in the ith valley. The contribution of the conduction-band electrons to the $\Gamma_{\vec{k}}^{\text{out}}$ is given by

$$\Gamma_{\vec{k}}^{\text{out}(c)} \simeq (5e^4 m^*/\pi\hbar^3\epsilon_\infty^2) I(k, K_-), \quad \text{(B7a)}$$

where

$$K_- = K - \tfrac{1}{2}\Delta, \quad \text{(B7b)}$$

$$I(k, K_-) = 4\ln\left(\frac{k + K_-}{k - K_-}\right) - 2\frac{K_-}{k} + \frac{2K_-(k^4 + K_-^4)(k^2 + \frac{1}{3}K_-^2)}{k(k^2 - K_-^2)^3} + \frac{8kK_-^2}{(k^2 - K_-^2)^2} + \frac{4K_-^3 - 8k^2K_-}{k(k^2 - K_-^2)}. \quad \text{(B7c)}$$

Let $k = K_- + \delta$. As $\delta \to 0$,

$$I(k, K_-) \simeq 4\ln(4K_-/\delta) + \tfrac{5}{8}(K_-/\delta)^3 + O(1/\delta^2) \to \infty. \quad \text{(B8)}$$

For $k = K + \frac{1}{2}\Delta$,

$$I \simeq 16 \text{ if } K = 2.5\Delta, \quad \text{(B9a)}$$

$$I \simeq 26 \text{ if } K = 2.9\Delta. \quad \text{(B9b)}$$

Thus, $\Gamma_{\vec{k}}^{\text{out}(c)} \geq 80e^4 m^*/\pi\hbar^3\epsilon_\infty^2 \simeq 3 \times 10^{14}$ sec^{-1}. The contribution of the holes is about the same; thus $\Gamma_{\vec{k}}^{\text{out}} \geq 6 \times 10^{14}$ sec^{-1} which is greater than the absorption rate. Finally, we note that when we include further screening effects this rate drops to around 10^{13} sec^{-1}.

APPENDIX C: RADIATION ABSORPTION RATES

(a) *Direct absorptions.* From the diagrams shown in Fig. 8, we readily find

$$\left.\frac{\partial N_{\lambda\vec{q}}^{(\mathrm{rad})}}{\partial t}\right|_{\mathrm{DA}} = -2\left(\frac{2\pi}{\hbar}\right)\left(\frac{e}{mc}\right)^2\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)N_{\lambda\vec{q}}^{(\mathrm{rad})}$$
$$\times\sum_{\vec{k}}|\langle c\Gamma,\vec{k}+\vec{q}|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|v,\vec{k}\rangle|^2\,[1-f_H(\vec{k})][1-f_{c\Gamma}(\vec{k}+\vec{q})]\delta(E_{c\Gamma}(\vec{k}+\vec{q})+E_H(\vec{k})-\hbar\omega_q)$$
$$+2\left(\frac{2\pi}{\hbar}\right)\left(\frac{e}{mc}\right)^2\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)(N_{\lambda\vec{q}}^{(\mathrm{rad})}+1)\sum_{\vec{k}}|\langle v,\vec{k}|e^{-i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|c\Gamma,\vec{k}+\vec{q}\rangle|^2 f_{c\Gamma}(\vec{k}+\vec{q})f_H(\vec{k})\delta(E_{c\Gamma}(\vec{k}+\vec{q})$$
$$+E_H(\vec{k})-\hbar\omega_q). \tag{C1}$$

Here, $\hat{\xi}_\lambda(\vec{q})$ denotes photon polarization. The meaning of the other symbols are as defined in the text. The sum $\sum_{\vec{k}}$ also includes spin summation. The extra factor of 2 in front of the terms is due to the fact that we have two degenerate valence bands. The photon momenta are completely negligible compared to the electron and hole momenta. Therefore, q's in the interband matrix element of $\vec{p}$, in f's and in the δ functions can be omitted. Also, the radiation is primarily at the circular frequency ω_0 and the corresponding $N^{(\mathrm{rad})}\gg 1$. Therefore, the spontaneous emission term can be neglected. To obtain

$$\left.\frac{\partial N}{\partial t}\right|_{\mathrm{DA}} = \frac{1}{N_0}\sum_{\lambda\vec{q}}\left.\frac{\partial N_{\lambda\vec{q}}^{(\mathrm{rad})}}{\partial t}\right|_{\mathrm{DA}},$$

we perform $\sum_{\lambda\vec{q}}$ on (C1), and find

$$\left.\frac{\partial N}{\partial t}\right|_{\mathrm{DA}} = -\left(\frac{8\pi^2e^2}{3m^2\epsilon_\infty\omega_0}\right)N(t)\int\frac{2d^3\vec{k}}{(2\pi)^3}|\langle c\Gamma,\vec{k}|\vec{p}|v,\vec{k}\rangle|^2[1-f_{c\Gamma}(\vec{k})-f_H(\vec{k})]\delta(E_{c\Gamma}(\vec{k})+E_H(\vec{k})-\hbar\omega_0). \tag{C2}$$

The interband matrix element of $\vec{p}$ near the Γ-valley band edge is given by

$$|\langle c\Gamma,\vec{k}|\vec{p}|v,\vec{k}\rangle|^2\simeq m^2E_0(m_0+m_h)/2m_0m_h. \tag{C3}$$

In our approximate band structure, the electron and hole energies are given by

$$E_{c\Gamma}(\vec{k})=E_0+\hbar^2k^2/2m_0,\quad E_H(\vec{k})=\hbar^2k^2/2m_h. \tag{C4}$$

The integral over the δ function is trivial, and yields

$$\left.\frac{\partial N}{\partial t}\right|_{\mathrm{DA}} = -N(t)\left(\frac{4e^2E_Q}{3\epsilon_\infty\hbar^2\omega_0}\right)\left(\frac{2m_Qm_h(\hbar\omega_Q-E_Q)}{\hbar^2(m_0+m_h)}\right)^{1/2}$$
$$\times\left(1-\left\{1+\exp\left[\beta\left(\frac{m_Q}{m_0+m_h}(\hbar\omega_0-E_0)+E_H\right)\right]\right\}^{-1}-\left\{1+\exp\left[\beta\left(E_0+\frac{m_h}{m_0+m_h}(\hbar\omega_0-E_0)-E_F\right)\right]\right\}^{-1}\right), \tag{C5}$$

$$\left.\frac{\partial N}{\partial t}\right|_{\mathrm{DA}} = -N(t)\alpha_0[1-F(z_4)-F(z_5)]$$
$$= -N(t)\alpha_D(t). \tag{28}$$

To find

$$\left.\frac{\partial n}{\partial t}\right|_{\mathrm{DA}} = \frac{1}{N_0}\int\frac{2d^3\vec{k}}{(2\pi)^3}\left.\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right|_{\mathrm{DA}},\quad \left.\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right|_{\mathrm{DA}}$$

can be calculated in completely analogous manner. It is given by the negative of the righthand side of (C1) if $\sum_{\vec{k}}$ is replaced by $\sum_{\lambda\vec{q}}$. Thus, $\partial n/\partial t|_{\mathrm{DA}}$ is given by the negative of the right-hand side of (C2), and

$$\left.\frac{\partial n}{\partial t}\right|_{\mathrm{DA}} = N(t)\alpha_0[1-F(z_4)-F(z_5)]$$
$$= N(t)\alpha_D(t). \tag{C6}$$

To find $\partial n/\partial t|_{\mathrm{DA}}$ we need $\partial f_H(\vec{k})/\partial t|_{\mathrm{DA}}$:

$$\left.\frac{\partial u}{\partial t}\right|_{\mathrm{DA}} = \frac{1}{N_0\hbar\omega_0}\left(\int\frac{2d^3\vec{k}}{(2\pi)^3}E_{c\Gamma}(\vec{k})\left.\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right|_{\mathrm{DA}}\right.$$
$$\left.+2\int\frac{2d^3\vec{k}}{(2\pi)^3}E_H(\vec{k})\left.\frac{\partial f_H(\vec{k})}{\partial t}\right|_{\mathrm{DA}}\right). \tag{16a}$$

Because of the two degenerate valence bands

$$\left.\frac{\partial f_H(\vec{k})}{\partial t}\right|_{\mathrm{DA}} = \frac{1}{2}\left.\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right|_{\mathrm{DA}}. \tag{C7}$$

Thus,

$$\left.\frac{\partial u}{\partial t}\right|_{\mathrm{DA}} = \frac{1}{N_0\hbar\omega_0}\int\frac{2d^3\vec{k}}{(2\pi)^3}[E_{c\Gamma}(\vec{k})+E_H(\vec{k})]\left.\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\right|_{\mathrm{DA}}.$$

Because of the δ function $\delta(E_{c\Gamma}(\vec{k}) + E_H(\vec{k}) - \hbar\omega_0)$ in $\partial f_{c\Gamma}(\vec{k})/\partial t\big|_{DA}$, $E_{c\Gamma}(\vec{k}) + E_H(\vec{k}) = \hbar\omega_0$, and

$$\frac{\partial n}{\partial t}\bigg|_{DA} = \frac{1}{N_0}\int \frac{2d^3\vec{k}}{(2\pi)^3}\frac{\partial f_{c\Gamma}(\vec{k})}{\partial t}\bigg|_{DA} = \frac{\partial n}{\partial t}\bigg|_{DA} = N(t)\alpha_D(t)\,. \tag{27a}$$

(b) *Free-carrier absorptions.* Let us consider the conduction band electrons first. Since the free-carrier absorptions strongly depend on density, and since the density of electrons in the Γ valley is negligible compared to the L-X valleys, we will consider the L-X valleys only, and ignore the Γ valley.

With the help of the diagrams, like the ones shown in Fig. 9, we find

$$\begin{aligned}
\frac{\partial N_{\lambda\vec{q}}^{(rad)}}{\partial t}\bigg|_{FCA\text{-}EL} = &-\left(\frac{2\pi}{\hbar}\right)\sum_{\vec{k}}\sum_{\vec{q}'}\sum_{\mu}\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)N_{\lambda\vec{q}}^{(rad)}\left(\frac{\hbar}{2\rho\Omega_{\mu q'}}\right)\left(\frac{e}{mc}\right)^2(1+\eta_{\mu q'}) \\
&\times f_{cj}(\vec{k})[1-f_{cj}(\vec{k}+\vec{q}-\vec{q}')]\delta(E_{cj}(\vec{k})+\hbar\omega_q - E_{cj}(\vec{k}+\vec{q}-\vec{q}') - \hbar\Omega_{\mu q'}) \\
&\times\Bigg|\frac{Q_\mu(\vec{q}')\langle cj,\vec{k}+\vec{q}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|cj,\vec{k}+\vec{q}\rangle\langle cj,\vec{k}+\vec{q}|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}\rangle}{E_{cj}(\vec{k})+\hbar\omega_q - E_{cj}(\vec{k}+\vec{q})} \\
&\quad + \frac{Q_\mu(\vec{q}')\langle cj,\vec{k}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|cj,\vec{k}\rangle\langle cj,\vec{k}+\vec{q}-\vec{q}'|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}-\vec{q}'\rangle}{E_{cj}(\vec{k}+\vec{q}-\vec{q}') - E_{cj}(\vec{k}-\vec{q}') - \hbar\omega_q}\Bigg|^2 \\
&-\left(\frac{2\pi}{\hbar}\right)\sum_{\vec{k}}\sum_{\vec{q}'}\sum_{\mu}\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)N_{\lambda\vec{q}}^{(rad)}\left(\frac{\hbar}{2\rho\Omega_{\mu q'}}\right)\left(\frac{e}{mc}\right)^2\eta_{\mu q'} \\
&\times f_{cj}(\vec{k})[1-f_{cj}(\vec{k}+\vec{q}+\vec{q}')]\delta(E_{cj}(\vec{k})+\hbar\omega_q - E_{cj}(\vec{k}+\vec{q}+\vec{q}') + \hbar\Omega_{\mu q'}) \\
&\times\Bigg|\frac{Q_\mu(\vec{q}')\langle cj,\vec{k}+\vec{q}+\vec{q}'|e^{i\vec{q}'\cdot\vec{r}}|cj,\vec{k}+\vec{q}\rangle\langle cj,\vec{k}+\vec{q}|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}\rangle}{E_{cj}(\vec{k})+\hbar\omega_q - E_{cj}(\vec{k}+\vec{q})} \\
&\quad + \frac{Q_\mu(\vec{q}')\langle cj,\vec{k}+\vec{q}'|e^{i\vec{q}'\cdot\vec{r}}|cj,\vec{k}\rangle\langle cj,\vec{k}+\vec{q}+\vec{q}'|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}+\vec{q}'\rangle}{E_{cj}(\vec{k}+\vec{q}+\vec{q}') - E_{cj}(\vec{k}+\vec{q}') - \hbar\omega_q}\Bigg|^2 \\
&+\left(\frac{2\pi}{\hbar}\right)\sum_{\vec{k}}\sum_{\vec{q}'}\sum_{\mu}\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)(N_{\lambda\vec{q}}^{(rad)}+1)\left(\frac{\hbar}{2\rho\Omega_{\mu q'}}\right)\left(\frac{e}{mc}\right)^2(1+\eta_{\mu q'}) \\
&\times f_{cj}(\vec{k})[1-f_{cj}(\vec{k}-\vec{q}-\vec{q}')]\delta(E_{cj}(\vec{k})-\hbar\omega_q - \hbar\Omega_{\mu q'} - E_{cj}(\vec{k}-\vec{q}-\vec{q}')) \\
&\times\Bigg|\frac{Q_\mu(\vec{q}')\langle cj,\vec{k}-\vec{q}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|cj,\vec{k}-\vec{q}\rangle\langle cj,\vec{k}-\vec{q}|e^{-i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}-\vec{q}'\rangle}{E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q})-\hbar\omega_q} \\
&\quad + \frac{Q_\mu(\vec{q}')\langle cj,\vec{k}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|cj,\vec{k}\rangle\langle cj,\vec{k}-\vec{q}-\vec{q}'|e^{-i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}-\vec{q}'\rangle}{E_{cj}(\vec{k}-\vec{q}-\vec{q}')+\hbar\omega_q - E_{cj}(\vec{k}-\vec{q}')}\Bigg|^2 \\
&+\left(\frac{2\pi}{\hbar}\right)\sum_{\vec{k}}\sum_{\vec{q}'}\sum_{\mu}\left(\frac{2\pi c^2\hbar}{\epsilon_\infty\omega_q}\right)(N_{\lambda\vec{q}}^{(rad)}+1)\left(\frac{\hbar}{2\rho\Omega_{\mu q'}}\right)\left(\frac{e}{mc}\right)^2\eta_{\mu q'} \\
&\times f_{cj}(\vec{k})[1-f_{cj}(\vec{k}+\vec{q}'-\vec{q})]\delta(E_{cj}(\vec{k})+\hbar\Omega_{\mu q'} - \hbar\omega_q - E_{cj}(\vec{k}+\vec{q}'-\vec{q})) \\
&\times\Bigg|\frac{Q_\mu(\vec{q}')\langle cj,\vec{k}+\vec{q}'-\vec{q}|e^{i\vec{q}'\cdot\vec{r}}|cj,\vec{k}-\vec{q}\rangle\langle cj,\vec{k}-\vec{q}|e^{i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}\rangle}{E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q})-\hbar\omega_q} \\
&\quad + \frac{Q_\mu(\vec{q}')\langle cj,\vec{k}+\vec{q}'|e^{i\vec{q}'\cdot\vec{r}}|cj,\vec{k}\rangle\langle cj,\vec{k}+\vec{q}'-\vec{q}|e^{-i\vec{q}\cdot\vec{r}}\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}|cj,\vec{k}+\vec{q}'\rangle}{E_{cj}(\vec{k}+\vec{q}'-\vec{q})-E_{cj}(\vec{k}+\vec{q}')+\hbar\omega_q}\Bigg|^2 .
\end{aligned} \tag{C8}$$

As before, j refers to the valleys, μ to the phonon modes. $\Omega_{\mu q'}$ and $\eta_{\mu q'}$ are phonon frequencies and phonon

distributions, respectively. $Q_\mu(\vec{q}')$ are the phonon-electron coupling parameters discussed in the text. We will assume that there are no phonon instabilities, and that

$$\eta_{\mu q'} = \left[\exp\left(\frac{\hbar\Omega_{\mu q'}}{k_B T_L}\right) - 1\right]^{-1} \simeq \frac{k_B T_L}{\hbar\Omega_{\mu q'}}. \tag{C9}$$

As before, the light momenta $q \ll k, q'$ and can be neglected in f's, δ functions, and the matrix elements. In the jth valley, for the Bloch states,[20]

$$\langle cj, \vec{k}+\vec{q}| e^{i\vec{q}\cdot\vec{r}} \hat{\xi}_\lambda(\vec{q})\cdot\vec{p}\,|cj,\vec{k}\rangle \simeq \langle cj,\vec{k}|\,\hat{\xi}_\lambda(\vec{q})\cdot\vec{p}\,|cj,\vec{k}\rangle \simeq \hbar(m/mc)\hat{\xi}_\lambda(\vec{q})\cdot\vec{k}. \tag{C10}$$

Since q' is small compared to the reciprocal-lattice vector[20]

$$\langle cj, k\pm q'|\, e^{\pm i\vec{q}'\cdot\vec{r}}\,|cj,k\rangle \simeq 1. \tag{C11}$$

Phonon energies $\hbar\Omega_{\mu q'} \ll E_{cj}$, $\hbar\omega_q$ and can be neglected in the δ functions. We will also neglect spontaneous photon emission terms, as done before in direct absorptions. With these approximations, and after a redefinition of integration variables, we obtain

$$\left.\frac{\partial N_{\lambda\vec{q}}}{\partial t}\right|_{\text{FCA-EL}} = -\left(\frac{2\pi e^2\hbar}{m_c^2\epsilon_\infty\omega_q^3\rho}\right) N_{\lambda\vec{q}} \sum_{j\vec{k}}\sum_{\mu}\sum_{\vec{q}'} \frac{|Q_\mu(\vec{q}')|^2}{\Omega_{\mu q'}} |\hat{\xi}_\lambda(\vec{q})\cdot\vec{q}'|^2 \left(1+2\frac{k_B T_L}{\hbar\Omega_{\mu q'}}\right)$$

$$\times [f_{cj}(\vec{k}) - f_{cj}(\vec{k}+\vec{q}')]\,\delta(E_{cj}(\vec{k}) + \hbar\omega_q - E_{cj}(\vec{k}+\vec{q}')). \tag{C12}$$

We again average over light polarization, then perform $(1/N_0)\sum_{\lambda\vec{q}}$, and find

$$\left.\frac{\partial N}{\partial t}\right|_{\text{FCA-EL}} = -N\left(\frac{2\pi^2 e^2\hbar}{3m_c^2\epsilon_\infty\omega_0^3\rho}\right)\sum_{j\vec{k}}\sum_{\mu}\sum_{\vec{q}'} \frac{|Q_\mu(\vec{q}')|^2 (q')^2}{\Omega_{\mu q'}} \left(1+2\frac{k_B T_L}{\hbar\Omega_{\mu q'}}\right)[f_{cj}(\vec{k}) - f_{cj}(\vec{k}+\vec{q}')]\,\delta(E_{cj}(\vec{k}) + \hbar\omega_0 - E_{cj}(\vec{k}+\vec{q}')). \tag{C13}$$

In our approximate band structure,

$$E_{cL,X}(\vec{k}) = E_G + \hbar^2(\vec{k}-\vec{k}_{L,X})^2/2m_c. \tag{2a}$$

Substituting this in (C13) and summing over the valleys, changing the integration variable $\vec{k}$, we find

$$\left.\frac{\partial N}{\partial t}\right|_{\text{FCA-EL}} = -N\left(\frac{20\pi^2 e^2\hbar}{3m_c^2\epsilon_\infty\omega_0^3\rho}\right)\int\frac{2d^3\vec{k}}{(2\pi^3)}\left[F\left(\frac{E_G+\hbar^2k^2/2m_c - E_F}{k_BT}\right) - F\left(\frac{E_G+\hbar^2k^2/2m_c+\hbar\omega_0 - E_F}{k_BT}\right)\right]$$

$$\times\sum_\mu\int\frac{d^3\vec{q}'}{(2\pi)^3}\,\frac{|Q_\mu(\vec{q}')|^2(q')^2}{\Omega_{\mu q'}}\left(1+2\frac{k_BT_L}{\hbar\Omega_{\mu q'}}\right)\delta\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0-\frac{\hbar^2(\vec{k}+\vec{q}')^2}{2m_c}\right), \tag{C14}$$

where $F(x) = (1+e^x)^{-1}$. For optical phonons,

$$\sum_\mu\int\frac{d^3\vec{q}'}{(2\pi)^3}\,\frac{|Q_\mu(\vec{q}')|^2(q')^2}{\Omega_{\mu q'}}\left(1+2\frac{k_BT_L}{\hbar\Omega_{\mu q'}}\right) - \left(1+2\frac{k_BT_L}{\hbar\Omega_0}\right)\left(\frac{Q_0^2}{\Omega_0}\right)\int\frac{d^3\vec{q}'}{(2\pi)^3}(q')^2.$$

Owing to the δ function, integration over $\vec{q}'$ is quite simple:

$$\int d^3\vec{q}'(q')^2\delta\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0-\frac{\hbar^2(\vec{k}+\vec{q}')^2}{2m_c}\right) = \frac{2\pi m_c^3}{\hbar^6 k}\left\{\left[\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0\right)^{1/2}+\left(\frac{\hbar^2k^2}{2m_c}\right)^{1/2}\right]^4 - \left[\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0\right)^{1/2}-\left(\frac{\hbar^2k^2}{2m_c}\right)^{1/2}\right]^4\right\}. \tag{C15}$$

For acoustical phonons, $\Omega_{\mu q'} \simeq c_A q'$ and

$$\sum_{\mu}\int \frac{d^3\vec{q}'}{(2\pi)^3}\,\frac{|Q_\mu(\vec{q}')|^2(q')^2}{\Omega_{\mu q'}}\left(1+2\frac{k_BT_L}{\hbar\Omega_{\mu q'}}\right)$$

$$-\int\frac{d^3q'}{(2\pi)^3}\,\frac{(\Lambda_d^2+\frac{2}{3}\Lambda_d\Lambda_u+\frac{1}{3}\Lambda_u^2)}{c_A q'}\,(q')^4\left(1+\frac{2k_BT_L}{\hbar c_A q'}\right)-\frac{1}{(2\pi)^3}\int d^3\vec{q}'\left(\frac{\Lambda^2}{c_A}\right)(q')^3\left(1+\frac{2k_BT_L}{\hbar c_A q'}\right).$$

One of the integrations is the same as in (C13). The other gives

$$d^3\vec{q}'\,(q')^3\delta\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0-\frac{\hbar^2(\vec{k}+\vec{q}')^2}{2m_c}\right)$$

$$=\left(\frac{2\pi m_c}{5\hbar^2}\right)\left(\frac{2m_c}{\hbar^2}\right)^{5/2}\left\{\left[\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0\right)^{1/2}+\left(\frac{\hbar^2k^2}{2m_c}\right)^{1/2}\right]^5-\left[\left(\frac{\hbar^2k^2}{2m_c}+\hbar\omega_0\right)^{1/2}-\left(\frac{\hbar^2k^2}{2m_c}\right)^{1/2}\right]^5\right\}. \quad \text{(C16)}$$

Substituting (C15) and (C16) in (C14), performing the remaining angular integrations (of $\vec{k}$), then letting $y=\hbar^2k^2/2\hbar\omega_0m_c$, we obtain

$$\left.\frac{\partial N}{\partial t}\right|_{\text{FCA-EL}}=-N\left\{\left[\left(\frac{5e^2m_c^2Q_0^2}{3\epsilon_\infty\rho\pi^2\hbar^4\Omega_0}\right)\left(1+\frac{2k_BT_L}{\hbar\Omega_0}\right)+\left(\frac{10\Lambda^2e^2m_c^2}{3\pi^2c_A^2\epsilon_\infty\rho\hbar^5}\right)(k_BT_L)\right]\int_0^\infty dy\,\{[(y+1)^{1/2}+(y)^{1/2}]^4\right.$$

$$-[(y+1)^{1/2}-(y)^{1/2}]^4\}\left[F\left(\frac{y+\delta_G-\epsilon}{\mathcal{T}}\right)-F\left(\frac{y+1+\delta_G-\epsilon}{\mathcal{T}}\right)\right]+\left(\frac{e^2\Lambda^2(2m_c)^{5/2}(\hbar\omega_0)^{1/2}}{3\pi^2\epsilon_\infty\rho c_A\hbar^5}\right)$$

$$\left.\times\int_0^\infty dy\,\{[(y+1)^{1/2}+(y)^{1/2}]^5-[(y+1)^{1/2}-(y)^{1/2}]^5\}\left[F\left(\frac{y+\delta_G-\epsilon}{\mathcal{T}}\right)-F\left(\frac{y+1+\delta_G-\epsilon}{\mathcal{T}}\right)\right]\right\}. \quad \text{(C17)}$$

Thus,

$$\left.\frac{\partial N}{\partial t}\right|_{\text{FCA-EL}}=-N\left[\alpha_1\times5\left(\frac{m_c}{m_h}\right)^2\left(1+2\frac{\omega_0}{\Omega_0}\,\mathcal{T}_L\right)\zeta_0(z_1)+\alpha_2\,\mathcal{T}_L\times5\left(\frac{m_c}{m_h}\right)^2\zeta_0(z_1)+\alpha_3\times5\left(\frac{m_c}{m_h}\right)^{5/2}\zeta_1(z_1)\right]. \quad \text{(C18)}$$

$\partial N/\partial t|_{\text{FCA-HOLES}}$ can be obtained from (C18) by simply omitting (m_c/m_h) factors, replacing $z_1=(\delta_G-\epsilon)/\mathcal{T}$ by $z_2=h/\mathcal{T}$, and dividing by 5. Adding electron and hole contributions we obtain $\partial N/\partial t|_{\text{FCA}}$ given by (39).

$\partial u/\partial t|_{\text{FCA}}$ can be obtained by a simple energy-conservation argument:

$$\left.\frac{\partial u}{\partial t}\right|_{\text{FCA}}=\frac{1}{\hbar\omega_0N_0}\left.\frac{\partial U}{\partial t}\right|_{\text{FCA}}=\frac{-1}{\hbar\omega_0N_0}\left.\frac{\partial(\hbar\omega_0N_0N)}{\partial t}\right|_{\text{FCA}}=\alpha_{\text{FCA}}(t)N(t). \quad \text{(C19)}$$

APPENDIX D: PHONON-ASSISTED RELAXATION

Phonon-assisted relaxation may be represented by the diagrams shown in Fig. 12. Let us consider the L-X valley electrons first. Using the electron-phonon coupling Hamiltonian given by (35), we find

$$\left.\frac{\partial f_{cj}(\vec{k})}{\partial t}\right|_{\text{REL}}=\frac{2\pi}{\hbar}\sum_\mu\sum_{\vec{q}'}\frac{|Q_\mu(\vec{q}')|^2\hbar}{2\rho\Omega_{\mu q'}}$$

$$\times[(\eta_{\mu q'}+1)\{|\langle cj,\vec{k}|e^{-i\vec{q}'\cdot\vec{r}}|cj,\vec{k}+\vec{q}'\rangle|^2f_{cj}(\vec{k}+\vec{q}')[1-f_{cj}(\vec{k})]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$

$$-|\langle cj,\vec{k}-\vec{q}'|e^{-i\vec{q}'\cdot\vec{r}}|cj,\vec{k}\rangle|^2f_{cj}(\vec{k})[1-f_{cj}(\vec{k}-\vec{q}')]\delta(E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'})\}$$

$$+\eta_{\mu q'}\{|\langle cj,\vec{k}|e^{i\vec{q}'\cdot\vec{r}}|cj,\vec{k}-\vec{q}'\rangle|^2f_{cj}(\vec{k}-\vec{q}')[1-f_{cj}(\vec{k})]\delta(E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'})$$

$$-|\langle cj,\vec{k}+\vec{q}'|e^{i\vec{q}'\cdot\vec{r}}|cj,\vec{k}\rangle|^2f_{cj}(\vec{k})[1-f_{cj}(\vec{k}+\vec{q}')]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})\}]. \quad \text{(D1)}$$

Using (C11), we find

$$\left.\frac{\partial f_{cj}(\vec{k})}{\partial t}\right|_{REL} = \frac{\pi}{\rho}\sum_{\mu}\sum_{\vec{q}'}\frac{|Q_{\mu}(\vec{q}')|^2}{\Omega_{\mu q'}}[\eta_{\mu q'}\{[f_{cj}(\vec{k}+\vec{q}')-f_{cj}(\vec{k})]\,\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$-[f_{cj}(\vec{k})-f_{cj}(\vec{k}-\vec{q}')]\delta(E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'})\}$$
$$+f_{cj}(\vec{k}+\vec{q}')[1-f_{cj}(\vec{k})]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$-f_{cj}(\vec{k})[1-f_{cj}(\vec{k}-\vec{q}')]\delta(E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'})]. \tag{D2}$$

We multiply (D2) by $E_{cj}(\vec{k})$ and sum over the electron states and the valleys:

$$\left.\frac{\partial u_g}{\partial t}\right|_{REL} = \frac{\pi}{\rho N_0\hbar\omega_0}\sum_{j\vec{k}}\sum_{\mu\vec{q}'}\frac{|Q_{\mu}(\vec{q}')|^2}{\Omega_{\mu q'}}[\eta_{\mu q'}E_{cj}(\vec{k})\{[f_{cj}(\vec{k}+\vec{q}')-f_{cj}(\vec{k})]\,\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$-[f_{cj}(\vec{k})-f_{cj}(\vec{k}-\vec{q}')]\delta(E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'})\}$$
$$+E_{cj}(\vec{k})f_{cj}(\vec{k}+\vec{q}')[1-f_{cj}(\vec{k})]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$-E_{cj}(\vec{k})f_{cj}(\vec{k})[1-f_{cj}(\vec{k}-\vec{q}')]\delta(E_{cj}(\vec{k})-E_{cj}(\vec{k}-\vec{q}')-\hbar\Omega_{\mu q'})]. \tag{D3}$$

Using the identities

$$f_{cj}(\vec{k}+\vec{q}')[1-f_{cj}(\vec{k})]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$=[(e^{\beta[E_{cj}(\vec{k}+\vec{q}')-E_F]}+1)^{-1}][1-(1+e^{\beta[E_{cj}(\vec{k})-E_F]})^{-1}]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$=(e^{\beta(\hbar\Omega_{\mu q'})}-1)^{-1}[(e^{\beta[E_{cj}(\vec{k})-E_F]}+1)^{-1}-(e^{\beta[E_{cj}(\vec{k})+\hbar\Omega_0-E_F]}+1)^{-1}]\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'})$$
$$=(e^{\beta(\hbar\Omega_{\mu q'})}-1)^{-1}[f_{cj}(\vec{k})-f_{cj}(\vec{k}+\vec{q}')\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_0), \tag{D4}$$

we find

$$\left.\frac{\partial u_g}{\partial t}\right|_{REL} = \left(\frac{\pi}{\rho N_0\hbar\omega_0}\right)\sum_{j\vec{k}}\sum_{\mu\vec{q}}\frac{|Q_{\mu}(\vec{q}')|^2}{\Omega_{\mu q'}}\left(\eta_{\mu q'}-\frac{1}{e^{\beta\hbar\Omega_{\mu q'}}-1}\right)$$
$$\times\{E_{cj}(k)[f_{cj}(k+q')-f_{cj}(k)]\,\delta(E_{cj}(k+q')-E_{cj}(k)-\hbar\Omega_{\mu q'})$$
$$-E_{cj}(k)[f_{cj}(k)-f_{cj}(k-q')]\,\delta(E_{cj}(k)-E_{cj}(k-q')-\hbar\Omega_{\mu q'})\}. \tag{D5}$$

Since phonon energies are small,

$$\eta_{\mu q'}-(e^{\beta\hbar\Omega_{\mu q'}}-1)^{-1}\simeq k_BT_L/\hbar\Omega_{\mu q'}-k_BT/\hbar\Omega_{\mu q'}. \tag{D6}$$

Letting $\vec{k}\to\vec{k}+\vec{q}'$ in the second term, we obtain

$$\left.\frac{\partial u_g}{\partial t}\right|_{REL} = \left(\frac{\pi}{\rho N_0\hbar\omega_0}\right)\sum_{j\vec{k}}\sum_{\mu\vec{q}}\frac{|Q_{\mu}(\vec{q}')|^2}{\Omega_{\mu q'}}\left(\frac{k_BT_L-k_BT}{\hbar\Omega_{\mu q'}}\right)[E_{cj}(\vec{k})-E_{cj}(\vec{k}+\vec{q}')][f_{cj}(\vec{k}+\vec{q}')-f_{cj}(\vec{k})]$$
$$\times\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'}). \tag{D7}$$

Owing to the δ function,

$$\left.\frac{\partial u_g}{\partial t}\right|_{REL} = \left(\frac{\pi}{\rho N_0\hbar\omega_0}\right)\sum_{j\vec{k}}\sum_{\mu\vec{q}}\frac{|Q_{\mu}(\vec{q}')|^2}{\Omega_{\mu q'}}(k_BT-k_BT_L)\,[f_{cj}(\vec{k}+\vec{q}')-f_{cj}(\vec{k})]\,\delta(E_{cj}(\vec{k}+\vec{q}')-E_{cj}(\vec{k})-\hbar\Omega_{\mu q'}). \tag{D8}$$

Redefining $\vec{k}$ in each valley, we obtain

$$\left.\frac{\partial u_g}{\partial t}\right|_{REL} = \frac{10(k_BT-k_BT_L)}{\rho N_0\hbar\omega_0(2\pi)^3}\sum_{\mu}\int d^3\vec{k}\int d^3\vec{q}'\,\frac{|Q_{\mu}(\vec{q}')|}{\Omega_{\mu q'}}$$
$$\times[f(E+E_G+\hbar\Omega_{\mu q'})-f(E+E_G)]\,\delta\left(\frac{\hbar^2kq'\cos\theta}{m_c}+\frac{\hbar^2q'^2}{2m_c}-\hbar\Omega_{\mu q'}\right), \tag{D9}$$

where $\cos\theta = \hat{k}\cdot\hat{q}'$ and

$$E = \hbar^2 k^2/2m_c\,, \quad f(E) = (e^{\beta(E-E_F)}+1)^{-1}\,. \tag{D10}$$

Since phonon energies are small,

$$f(E+E_G+\hbar\Omega_{\mu q'}) - f(E+E_G) \simeq \hbar\Omega_{\mu q'}\left(\frac{\partial f(E+E_G)}{\partial E}\right) \tag{D11}$$

and $\hbar\Omega_{\mu q'}$ can be omitted from the δ function. Thus,

$$\left.\frac{\partial u_g}{\partial t}\right|_{\rm REL} = \frac{10(k_B T - k_B T_L)}{\rho N_0 \omega_0 (2\pi)^5} \sum_\mu \int d^3\vec{k}' \int d^3\vec{q}'\, |Q_\mu(\vec{q}')|^2 \frac{\partial f(E+EG)}{\partial E}\, \delta\left(\frac{\hbar^2 k q' \cos\theta}{m_c} + \frac{\hbar^2 q'^2}{2m_c}\right). \tag{D12}$$

Substituting the properly angular averaged electron-phonon coupling coefficients we find

$$\left.\frac{\partial u_g}{\partial t}\right|_{\rm REL} = \frac{10(k_B T - k_B T_L)}{\rho N_0 \omega_0 (2\pi)^5} \int d^3\vec{k}' \int d^3\vec{q}'\, (Q_0^2 + \Lambda^2 q'^2) \frac{\partial f(E+E_G)}{\partial E}\, \delta\left(\frac{\hbar^2 k q' \cos\theta}{m_c} + \frac{\hbar^2 q'^2}{2m_c}\right). \tag{D13}$$

The remaining integrals are trivial, and we find (in terms of the normalized quantities)

$$\left.\frac{\partial u_g}{\partial t}\right|_{\rm REL} = -\Gamma^{(0)} 5 \left(\frac{m_c}{m_h}\right)^3 \mathcal{T}(\mathcal{T} - \mathcal{T}_L) \ln(1 + e^{-z_1}) - \Gamma^{(a)} 5 \left(\frac{m_c}{m_h}\right)^4 \mathcal{T}^2 (\mathcal{T} - \mathcal{T}_L) \kappa(z_1)\,. \tag{D14}$$

$\partial u_h/\partial t\,|_{\rm REL}$ can be obtained from (D14) by replacing z_1 by z_2, dividing by 5, and omitting (m_c/m_h) factors.

APPENDIX E: EXPERIMENTAL METHOD AND DETAILS

(a) *Introduction.* This appendix contains an elaboration of the experimental method. This presentation of the experimental apparatus is given in detail because picosecond pulse techniques are not a standard tool of solid-state physicists, and some aspects of sample preparation may be of interest to the "nonsolid staters." The first experimental section contains a description of the mode-locked Nd:glass laser system, electro-optical switch-out apparatus, and the pulse splitter and delay optics. The second section presents properties of the germanium sample, including purity, surface preparation, and thickness measurement. The third section discusses the detection system, and the fourth discusses possible sources of experimental error. Figure 31 shows the entire apparatus in block diagram form.

(b) *Pulse preparation.* The production of very short optical pulses by mode-locking requires a laser system with an active medium that provides a large gain bandwidth for light amplification. Nd^{+3}-glass, with a fluorescent bandwidth of 2.7×10^{12} Hz, has the potential of producing pulses as short as 0.3 psec. In practice, pulses between 5 and 15 psec duration are obtained. This deviation from the theoretically calculated pulse width is caused by the broadening of short pulses in a dispersive and nonlinear medium and by spectral narrowing due to preferential amplification of frequency components close to the gain maximum. In fact, pulse widths would be wider than those observed if it were not for the pulse shortening effects of the nonlinear saturable absorber that is used to passively mode lock the laser. For a discussion of the basic principles of mode-locking and the production of ultrashort laser pulses, the reader is referred to von der Linde.[1]

The mode-locked output from the Nd:glass laser system employed in these experiments consists of a series of approximately 100 pulses separated one from the next by approximately 8 nsec. Each pulse is typically 5–10 psec in duration and has an energy of 0.1 mJ at a wavelength of 1.06 μm. The laser produces single well mode-locked pulse trains on more than 80% of its firings. A brief description of the laser hardware follows.

The active lasing medium, a Brewster-cut Owens-Illinois ED-2S Nd:glass rod approximately 19 cm long and 6.4 mm in diameter, is housed in a Korad K1 laser head. The glass rod is pumped by a helical xenon flash lamp with an arc length of 28 in. A high-voltage, symmetrically shaped pulse

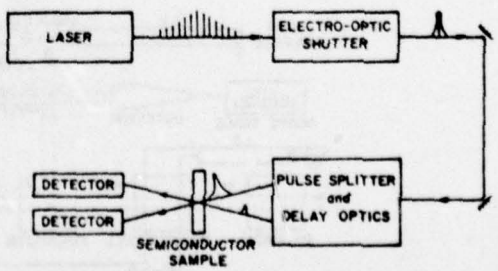


FIG. 31. Block diagram of the experiment.

from a 10-kV power supply is used to ionize the xenon gas. The high-voltage pulse is 0.5 msec in duration, and it typically delivers an energy of 600 J at 4.5 kV. Both flash lamp and glass rod are cooled by circulating water maintained at room temperature. The power supply consists of a 60-μF capacitor bank plus pulse shaping electronics. The flash lamp and power supply are fired approximately once per minute.

The laser oscillator cavity is 120 cm long with a spherical mirror of radius 2.5 m with a reflectivity of 99.7% at 1.06 μm at one end and a flat mirror with a reflectivity of 27% at the output end. The laser is mode-locked by placing a 1-mm-thick cell containing at 10:1 solution of dichloroethane and Eastman Q-switch solution A9860 in contact with the output mirror. The output mirror forms one face of the dye cell. Thus it must be resistant to the solvent dichloroethane as well as withstand the high-power levels of the optical pulses. The mode-locking dye is replenished after each laser firing by siphoning fresh dye solution from a reservior with a hypodermic needle.

The single most critical component of the entire laser system is the glass laser rod. Self-focusing of the pulse propagating in the glass rod, due to the nonlinear refractive index n_2, is a major problem in achieving reliable mode-locking in this laser. Using Owens-Illinois type ED-2S laser glass, we obtained clean pulse trains on approximately 80% of the firings.

In order to perform the two experiments, a single pulse must be selected from the train of picosecond pulses. Figure 32 shows the electro-optical shutter used to perform this function. The pulse train passes through a Pockels cell placed between two Glan prisms which act as nonabsorbing crossed polarizers. If no voltage is applied to the Pockels cell, the light passes through with no change in polarization and is rejected by the second crossed polarizer. When the half wave voltage is applied to the Pockels cell, the light polarization is rotated by 90° and thus the light is transmitted by the second polarizer. Voltage is held off the Pockels cell until

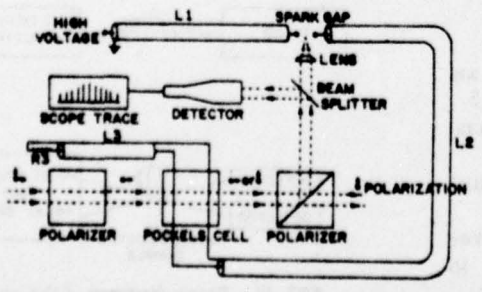


FIG. 32. Electro-optical shutter.

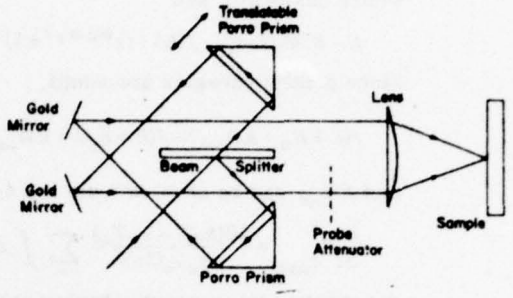


FIG. 33. Pulse splitter and delay optics.

the first part of the pulse train, which is reflected by the second Glan prism, breaks down a spark gap. Upon breakdown, a voltage pulse from a charged transmission line is applied to the Pockels cell; the length of the charged line is chosen such that the duration of the voltage pulse is equal to the temporal pulse separation in the train in order that one and only one pulse is switched out of the train.

In order to perform the two pulse or probe experiments, the single pulse selected from the train must be split in two and a variable delay must be introduced between these two pulses. Figure 33 is a detail of the pulse splitter and delay optics. Although a weak probe pulse is generally desired for the probe experiment, the pulse is split by a 50–50 beam splitter for ease in alignment of the two pulses coincident on the sample surface. The variable optical delay is introduced by means of a translatable Porro prism. The plano-convex lens, focal length $f = 20$ cm, serves to increase the optical energy density of the pulse and to angularly separate the two pulses for detection. The lens is placed after the pulse splitter and delay optics in order that the probe pulse spot size in the plane of the sample remains independent of the prism position. Alignment of the optical components and translation of the prism are critical in order that the two pulses are spatially coincident on the plane of the sample and remain that way during the experiment. Alignment of the two pulses is confirmed by inspection of burn patterns on exposed Polaroid film placed in the plane of the sample, because the red beams used for alignment of the laser are not coincident with the infrared laser output after passing through all the switch-out and pulse splitter and delay optics.

(c) *Crystal sample*. The semiconductor material used in these experiments is intrinsic germanium. The minimum resistivity of the high-purity single crystal sample is 40.0 Ω cm at room temperature. The sample was prepared from a crystal wafer 1 in. in diameter. First one surface was polished flat and smooth and then etched with Syton. The purpose of etching is to remove the damaged (from

mechanical polishing) surface layer and clean the surface of impurities in order that the optical measurements reflect the bulk properties of the solid. As is well known, etching the sample surface has an important effect on the observation of recombination radiation of germanium excited by a thermal source. After the first surface was prepared, the sample was bonded to a KZF-2 glass substrate, which has approximately the same coefficient of thermal expansion as germanium and is transparent at the laser wavelength. Then the sample was ground down to the desired thickness and the second surface was also polished and etched. Sample thickness is determined to be 5.2 μm by the interference fringe spacing of the reflection spectra measured in the infrared region of 3–15 μm. Crystallinity of the sample is verified by observing the x-ray diffraction pattern of the sample. During the measurements the sample on the substrate is mounted in an evacuated Dewar with quartz windows in order that the sample can be cooled. The sample temperature is measured by a thermocouple in contact with the crystal surface.

(d) *Detection system.* Two different types of detectors are used in these experiments. The entire pulse train (minus the switched out pulse) is monitored in order to verify that the pulse train is well mode-locked. To do this, the laser output is displayed on an ITT model F4000 biplanar vacuum photodiode and a Tektronix model 519 oscilloscope; the combined response time is approximately 1 nsec.

The primary measurement in these experiments is the measurement of energy in a single (incident, reflected, or transmitted) picosecond pulse. This is done by means of United Detector Technology model 6D silicon *p-i-n* photodiode operated in the photoconductive mode. When the laser pulse is incident on the large area photodiode, electron-hole pairs created in the intrinsic region junction are swept out by the reverse bias and charge an integrating capacitor. The voltage on this capacitor is proportional to the energy in the pulse and is read by means of a peak-detector-and-hold circuit and displayed on a digital voltmeter. The optical collection system is designed such that the detectors intercept radiation only from a small solid angle completely filled by the sample. Electrical shielding reduces pickup of electrical noise from the flash discharge. Detection of recombination radiation from germanium is eliminated by the sharp cutoff in spectral sensitivity of the silicon photodiodes at 1.1 μm.

The pulse energy detectors were calibrated both absolutely and relatively. Absolute calibration was performed against a calibrated Quantronix model 500 energy receiver and model 504 energy/power meter. Relative sensitivities of the detector units were measured with the units placed in their respective experimental configurations.

(e) *Experimental error.* We believe the main sources of error for both excitation and excitation-probe measurements to be the variation in pulse width and in energy from one laser firing to the next. The laser pulse train envelope is somewhat irreproducible, and the pulse width and shape vary with the location of the pulse in the train of pulses, as described by von der Linde.[1] This irreproducible and random nature of the pulse evolution within the laser cavity precludes the continued selection of identical pulses by the laser-triggered spark gap. The pulse width (unmeasured) is believed to range from 5 to 10 psec in duration. The dependence of the excitation-probe data on the level of excitation is clearly depicted in Fig. 4; the theoretical dependence of the single pulse (excitation transmission data on pulse width are illustrated in Fig. 25).

In addition to the uncertainty in pulse energy and width from data point to data point, the transverse mode structure of the laser is also uncertain (uncontrolled). Deviations of the transverse mode structure from the TEM_{00} mode will lead to "hot" spots on the surface of the semiconductor sample when the beam is focused. Variations in the positions of these "hot" spots as the mode structure changes from one firing to the next will result in variations in the degree of overlap between the excitation and probe beams and consequently will result in scatter in the probe transmission.

---

*During the course of this work we have appreciated the support of the ERDA, the NSF, the Alfred P. Sloan Foundation, the U. S. Air Force, (SAMSO), the Office of Naval Research, and the Al-Hazen Research Institute (Baghdad, Iraq).

[1]D. von der Linde, Appl. Phys. 2, 281 (1973).

[2]J. W. Shelton and J. A. Armstrong, IEEE J. Quantum Electron. QE-3, 696 (1967).

[3]C. J. Kennedy, J. C. Matter, A. L. Smirl, H. Weichel, F. A. Hopf, and S. V. Pappu, Phys. Rev. Lett. 32, 419 (1974).

[4]D. H. Auston and C. V. Shank, Phys. Rev. Lett. 32, 1120 (1974).

[5]A. L. Smirl, J. C. Matter, A. Elci, and M. O. Scully, Opt. Commun. 16, 118 (1976).

[6]D. von der Linde, O. Bernechker, and A. Laubereau, Opt. Commun. 2, 215 (1970).

[7]W. C. Dash and R. Newman, Phys. Rev. 99, 1151

(1955).

[8]H. Shih, M. Scully, W. H. Louisell, and W. B. McKnight, Appl. Opt. 12, 2198 (1973).

[9]M. Neuberger, *Handbook of Electronic Materials* (Plenum, New York, 1971), Vol. 5.

[10]Since the minima of the L valleys are exactly at the Brillouin-zone boundary, there are four of them instead of eight. But the minima of the valleys along Δ_1 are at $k \simeq 0.8kX$; therefore, there are six of them.

[11]Density-of-states effective mass is defined as $(m_t^2 m_l)^{1/3}$, where m_t and m_l are transverse and longitudinal effective masses, respectively.

[12]Indirect transitions are discussed in J. Bardeen, F. J. Blatt and L. H. Hall, in *Proceedings of the Conference on Photoconductivity, Atlantic City, November 4—6*, edited by R. G. Breckneridge *et al.* (Wiley, New York, 1956).

[13]A few notational comments: the tilda on a summation sign means that the summation is divided by the normalization density N_0. Electron or hole spin summations are not shown explicitly, but should be understood. The script letters denote the normalized quantities, and the corresponding ordinary type letters denote the corresponding unnormalized quantities in ordinary units (cgs).

[14]H. Y. Fan, M. L. Shepherd, and W. Spitzer, in Ref. 12.

[15]D. L. Dexter, in Ref. 12.

[16]A. H. Kahn, Phys. Rev. 97, 1647 (1955).

[17]H. Y. Fan, W. Spitzer, and R. J. Collins, Phys. Rev. 101, 566 (1956).

[18]R. Rosenberg and M. Lax, Phys. Rev. 112, 843 (1958).

[19]M. Lax and J. J. Hopfield, Phys. Rev. 124, 115 (1961).

[20]H. J. G. Meyer, Phys. Rev. 112, 298 (1958); J. Phys. Chem. Solids 8, 264 (1959).

[21]J. M. Lutthinger and W. Kohn, Phys. Rev. 97, 869 (1955).

[22]If there were no valleys, this statement would not be true. In that case the free-carrier absorptions could become dominant at high light intensities.

[23]C. A. Hogarth, in *Materials Used in Semiconductor Devices*, edited by C. A. Hogarth (Interscience, New York, 1965).

[24]Walter A. Harrison, Phys. Rev. 104, 1281 (1956).

[25]C. Herring and E. Vogt, Phys. Rev. 101, 944 (1956).

[26]B. N. Brockhouse and P. K. Iyengar, Phys. Rev. 111, 747 (1958).

[27]C. Herring, Bell Syst. Tech. J. 34, 237 (1955).

[28]$c_L = 2.31$ for Ge.

[29]P. A. Wolff, Phys. Rev. Lett. 24, 266 (1970).

[30]P. Nozières and D. Pines, Phys. Rev. 109, 741 (1958).

[31]J. J. Quinn and R. A. Ferrell, Phys. Rev. 112, 812 (1958).

[32]The portion of the incident pulse intensity that is reflected from the front surface must be subtracted from the incident pulse intensity to obtain the source term.

[33]The loss term takes into account the internal reflection from the rear surface of the crystal, which decreases transmission.

[34]The maximum refractive index change observed is $\Delta n/n \simeq -0.05$ (see Ref. 4).

[35]In the initial phase, practically all photons that are absorbed are used in increasing the electron density. Also, at this stage, the electron density is quite small, and therefore the high-temperature approximation can be made to calculate the electron and hole energy density. This energy density is then given by $nN_0(E_G + \frac{3}{2}k_BT) + nN_0 \times \frac{3}{2}k_BT$. Setting this equal to the radiation energy density that is absorbed, $nN_0\hbar\omega_0$, we find $k_BT = \frac{1}{5}(\hbar\omega_0 - E_G)$, or in terms of the normalized quantities, $\mathcal{T} = \frac{1}{5}(1 - \mathcal{E}_G)$.

[36]The authors would like to comment at this point that the absolute values of the probe transmissions shown in these figures are higher than those experimentally observed. This is a result of the approximations made in describing the Ge band structure. Further study has shown that the absolute transmission may be adjusted by removing some of these approximations. These points are discussed in a following paper.

[37]The refractive index for the optical frequency ω_0 is approximately given by $[\epsilon_\infty(1 - \omega_p^2/\omega_0^2)]^{1/2}$.

APPENDIX B

PHYSICS OF ULTRAFAST PHENOMENA IN SOLID-STATE PLASMAS

Accepted for publication in
Solid State Electronics

PHYSICS OF ULTRAFAST PHENOMENA IN SOLID STATE PLASMAS *

by

Ahmet Elci and Arthur L. Smirl
Physics Department
North Texas State University
Denton, Texas 76203

and

C. Y. Leung and Marlan O. Scully
Department of Physics and Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

ABSTRACT

Over the past half-decade measurements of the nonlinear, nonequilibrium optical properties of the germanium solid-state plasma have been carried out in several laboratories. These measurements involve, for example, the ultrafast relaxation of optically-excited electron-hole distributions in semiconductors and the photoluminescence spectrum of Ge at high excitation intensities. Here, we review our present theoretical understanding of these experiments and we discuss the theoretical limitations. We also extend our previous model, through a simple calculation, to include carrier diffusion.

I. INTRODUCTION

The development of lasers that can deliver high intensity optical pulses of extremely short duration has renewed and stimulated the interest in the measurement of the optical properties of semiconductors. Such measurements give information about the ultrafast dynamics of photogenerated electron-hole plasmas, as well as the band structure of the semiconductors. In this paper, we review some of our experiments in which picosecond optical pulses were used, their results, and our theoretical work concerning them. We also present the results of a simple calculation that illustrates the effect of including diffusion in our previous model. Other major work in this area, performed at Bell Laboratories, will be reviewed by D. H. Auston in an invited paper elsewhere in this issue.

In this paper, we will separate our discussion into two parts. The first part (Section II) describes the photogeneration of the carrier plasma by an intense ultrashort excitation pulse and the evolution of that plasma during the time this excitation pulse is incident on the sample. The second part (Section III) describes the temporal evolution of that photogenerated plasma after the excitation pulse has passed, as monitored by a weak probe pulse. In addition, we also present, in Section III, a simple analysis of the possible effects of carrier diffusion on these optical measurements.

II. OPTICAL GENERATION OF HIGH DENSITY, HOT ELECTRON HOLE PLASMAS

In this section we review experiments in which the electron-hole plasma is generated by the absorption of a single, ultrashort optical pulse in a thin semiconductor wafer. The transmission of this single pulse through the sample provides information concerning the generation and evolution of the plasma during the period the pulse is present in the semiconductor.

The nonlinear transmissions of thin germanium wafers have been measured by Kennedy et al.[1] and Smirl et al.[2] using ultrashort (5-10 psec) pulses at 1.06μm from a mode-locked Nd:glass laser. Figure 1 shows the transmission of a 5.2μm-thick Ge sample as a function of incident pulse energy for two sample temperatures, 77K and 300K.[2] These curves display two regions. There is a region in which the transmission is nearly constant and there is a nonlinear absorption region in which the transmission rises slowly and appears to peak where the crystal is damaged by the optical pulse. Similar measurements have been performed on thin samples of $Hg_{1-X}Cd_XTe$ by Matter et al.[3], as shown in Fig. 2. We shall concentrate on the Ge results, since the theoretical model for Ge is more nearly complete.

Recently, Elci et al.[4] have presented a theoretical model that accounts for the transmission of single optical pulses through thin Ge samples in terms of the energy band structure and other well-known parameters. The band structure of Ge is shown in Fig. 3. The data presented in Fig. 1 can be accounted

for in terms of this model in the following way. When the optical pulse enters the Ge crystal, it is absorbed by direct transitions. Thus, it creates a large number of electrons (holes) in the central valley of the conduction (valence) band. The electrons located in the central valley of the conduction band are rapidly ($\lesssim 10^{-14}$ sec) scattered to the conduction band side valleys by long wave vector phonons. Since these electrons are emptied from the central valley to side valleys at a rate that is comparable to the direct absorption rate, any decrease in the number of states available in the central valley for direct absorption is ultimately determined by a buildup of the populations in the side valleys. Carrier-carrier scattering events, which also occur at a rate comparable to the direct absorption rate, ensure that the electron and hole distributions will be Fermi-like. They also ensure that the Fermi distribution for holes and the Fermi distribution for electrons will reach a common temperature that is in general greater than the lattice temperature. Since the photon energy $\hbar\omega_o$ is greater than either the direct band gap E_o or the indirect gap E_G, a direct absorption event followed by phonon-assisted scattering of the electron from the central to side conduction band valleys results in an excess energy $\hbar\omega_o - E_G$ being given to thermal agitation. This excess energy results in a carrier temperature due to direct absorption that is much higher than the lattice temperature. Thus, direct absorption results in the creation of a large number of electrons (holes) in the conduction (valence) band with a high distribution temperature.

If the above mentioned processes were the only ones included in the model, the transmission of the thin Ge wafer would begin at its Beer's law value and rise as a function of incident pulse energy because of the saturation of the optically coupled states as a result of direct absorption. The observed rise in transmission as a function of incident pulse energy is, however, much slower than the band filling due to direct absorption would predict. The observed slower rise is caused by many body effects in the electron-hole plasma, as will be discussed below.

As a result of direct absorption, the conduction (valence) band of the sample contains a large number of electrons (holes) with a high distribution temperature. Electrons (holes) located high (low) in the conduction (valence) band can relax within that band by phonon emission. The effect of this relaxation is to reduce the carrier temperature and increase the lattice temperature. As the distribution cools by means of these intraband phonon-assisted transitions, carriers fill the states needed for absorption. Thus, phonon-assisted relaxation tends to increase, or further enhance, the sample transmission. Phonon-assisted relaxation occurs on a time scale comparable to the optical pulse width.

Many body effects in the electron-hole plasma, which become significant as the density of the carriers created by the optical pulse increases, cause the experimentally-measured transmission as a function of incident optical pulse energy

to rise more slowly than predicted by band filling caused by direct absorption alone. In particular, an electron or a hole moves in the screened coulomb field of other electrons and holes. The quantized long range part of the screened coulomb field is a plasmon. As the carrier density increases, because of direct absorption, the plasma frequency of the carriers increases. If the carrier density is sufficiently high, an electron in the central conduction band valley can recombine with a hole near the top of the valence bands while emitting a plasmon. Normally, plasmon-assisted recombination (PAR) can occur only if the plasma frequency ω_p is larger than the direct band gap frequency $E_o/\hbar$. However, in our case, owing to the strong perturbation of the solid by the optical pulse, the plasmon resonance is broadened considerably. Therefore, PAR can occur at plasma frequencies below $E_o/\hbar$. These collective oscillations have extremely short lifetimes and rapidly decay, transferring their energy to electrons and holes, thus increasing the carrier temperature. Therefore, as the carrier number increases, the PAR rate increases,which retards the growth of the carrier number and raises the carrier temperature. Thus, further increase in the sample transmission is slowed as the electrons (holes) are heated and removed from the optically coupled states. Free carrier absorption, although less effective, also serves to increase the distribution temperature while the pulse is incident on the sample.

Thus, we see that, physically, the single pulse transmission of Ge as a function of incident pulse energy (see Fig. 1) can be accounted for by direct interband transitions followed by heating of the carrier distribution. The solid lines in Fig. 1 represent the calculated transmissions using the model of Elci et al.[4] The agreement between theory and experiment is quite good.

Plasmons play an important role in determining the transmission properties of thin Ge wafers on picosecond time scales. Luckily, there is a property of PAR that makes it possible to test experimentally. PAR is sensitive to $\hbar\omega_p/E_o$. If the direct band gap E_o could be increased slightly while keeping the carrier density relatively constant, the PAR rate should decrease appreciably. This would result in a decreased plasma temperature for a given pulse energy, and one would expect the saturated transmission to increase owing to the carriers being located lower in the band. It is possible to increase E_o by subjecting Ge to pressure. Van Driel et al.[5] have measured the nonlinear transmission of picosecond optical pulses (1.06μm) through a thin Ge sample as a function of hydrostatic pressure up to 24 kbars, as shown in Fig. 4. At high intensities, the degree of enhanced transmission is observed to increase then decrease as the band gap is tuned with pressure. In terms of our model, the increase and then decrease in enhanced transmission with increasing pressure is caused by a decrease in the PAR rate, plasma temperature, and carrier density. The initial increase in the degree of enhanced transmission as the pressure is increased from 1 bar is primarily a result of a decrease in the

PAR rate caused by an increase in the direct gap E_o, which allows for a slightly more dense but cooler plasma. This results in the optically coupled electron-hole states being more completely occupied than at the lower pressure and, therefore, causes a higher transmission. This trend continues up to a pressure of approximately 10 kbar. Above this pressure, other effects begin to dominate. For example, the direct absorption coefficient in the Beer's law region will decrease with increasing pressure, reflecting a decrease in the joint density of states, for $\hbar\omega_o$ = 1.17 eV, as the valence and conduction bands move apart. This decrease in absorption coefficient results in a decrease in the carrier density and, thus, the plasma frequency. A more subtle effect in the reduction of the efficiency of the PAR process occurs because the indirect energy gaps do not increase as fast as the direct gap with increasing pressure. Therefore, the ratio of the central to side valley electron populations decreases with increasing pressure. Since the efficiency of the PAR process depends on the number of carriers in the central valley of the conduction band, the number of PAR events decreases as well.

Because of the role of the conduction band side valleys in the present interpretation of the single pulse transmission data, van Driel[6] et al. have measured the photoluminesence spectra of Ge as a function of excitation level at 1.06μm. Figure 5 shows the results such experiments performed at sample temperatures of 295 and 174K. Four different excitation irradiances (50,

500, 10^5 and 10^8 W/cm^2) were used at each temperature. The data at both temperatures show a peak at 1.76μm corresponding to the indirect gap recombination. In addition, there is a second peak at 1.5μm, corresponding to direct gap transitions. The direct gap photoluminescence intensity increases with increasing excitation level. The main features of the data can be understood by a simple argument. Recall that electron scattering from the central to side conduction band valleys by long wave vector phonons is extremely rapid. Thus, any buildup of the central valley electron population must be accompanied by a buildup of the populations in the ten side valleys. At low excitation levels, the number of electrons created is so small that an insignificant number exists in the central Γ valley; nearly all recombination is indirect. At higher excitation levels, the number of electrons in the central valley increases. Although this number is always much less than the number in the side valleys, the recombination rate across the direct gap far exceeds that across the indirect gap because the latter requires phonon participation. It is therefore possible with increasing carrier density to observe more direct than indirect gap luminescence. The theory, shown by solid lines in Fig. 5 , is seen to be in good agreement with experiment.

We now make a few additional comments concerning certain assumptions made in this model. An important assumption made in our calculations[4] is that of local thermodynamic equilibrium.

Although we have used the rapid rate for carrier-carrier collisions as a justification for taking the carrier distribution functionsto be Fermi-Dirac, it is none-the-less an assumption. The implications of this assumption are currently being investigated. If the distributions are assumed to be Fermi-Dirac, the calculations are greatly simplified. The plasma can then be described by a local temperature and Fermi energies. To this date, good results have been obtained with this assumption, as the experiments described above illustrate.

A more drastic assumption is to neglect the spatial variation of the parameters (temperature and Fermi energies) that characterize the electron-hole plasma throughout the interaction volume of the thin (approximately 5μm-thick) Ge wafer. Thus, in the interaction volume, electrons and holes are described by Fermi-Dirac distributions with appropriate Fermi energies for electrons and holes and a single temperature for both, all of which are independent of position. Therefore, the effects of the longitudinal diffusion of the carriers and of the optical pulse propagation are neglected. This is a severe limitation of the model and can be relaxed by allowing the plasma parameters to be dependent on both space and time. We include, in a simplistic manner, these spatial effects in the model for the first time in the next section of this paper. Other approximations and simplifications to the model are discussed by Latham et al.[7] in another paper in this issue.

III. EVOLUTION OF THE PHOTOGENERATED PLASMA

The electron-hole plasma created by an intense picosecond optical pulse evolves rapidly after the passage of that pulse, owing to the rapid relaxation processes that take place in semiconductors. Substantial information concerning this temporal evolution can be obtained by measuring the transmission of a second, weak picosecond pulse that is delayed in time with respect to the excitation pulse. In this section, we briefly review the experiments that monitor the temporal evolution of the hot plasma after the passage of the excitation pulse and our present understanding of these experiments. In addition, through a simple calculation, we consider the effect of diffusion on the present model.

The relative transmission of a probe pulse as a function of time delay after an excitation pulse for Ge sample temperatures of 77K and 297K is shown in Fig. 6 (from Ref. 2). In these experiments, the probe pulse had 5% of the energy of the excitation pulse. The excitation pulse was in the region of nonlinear transmission, and the probe was in the region of linear transmission (see Fig. 1). Note that the probe transmission rises rapidly, peaks, and then turns over for delays of the order of 100 psec. The solid lines in Fig. 6 are the probe transmissions calculated from the model of Elci et al.[4], in which only the intraband phonon-assisted carrier relaxation is taken into account; diffusion and other possible interactions are neglected. In view of these approximations, the agreement achieved is good.

The probe transmission can be understood in terms of the model of Elci et al.[4] in the following way. After the passage of the excitation pulse, the interaction region of the sample contains a large number of carriers with a high distribution temperature. The electrons (holes) are located high (low) in the bands because of the high distribution temperature, leaving the states that are optically coupled available for absorption. Thus, the probe transmission is small. Later, as the distribution cools by means of intraband phonon-assisted transitions and carriers fill the states needed for absorption, the probe transmission increases. Note that the PAR process is essentially turned off when the excitation pulse has passed through the sample.

One might plausibly attribute the decay of the probe transmission for long delays (Fig. 6) to carrier diffusion.[8,9] Because of the large focused spot size for the optical pulse (250μm in diameter), diffusion transverse to the direction of light propagation is entirely neglible. However, diffusion of carriers from a region near the sample surface into the bulk crystal in the direction of light propagation can be significant. We term the diffusion in the direction of light propagation longitudinal. Reflectivity experiments that measure the longitudinal diffusion of an optically-created surface carrier density into the crystal bulk have been carried out by Auston and Shank.[8] However, note that longitudinal carrier diffusion will not necessarily affect surface and bulk measurements in the same manner. When a probe pulse

is transmitted through a thin sample, it sees the same total number of carriers (neglecting Auger and other recombination) for all relevant delays; however, the carriers will be spatially rearranged because of diffusion. To investigate how the carrier diffusion affects the bulk transmission measurements, we perform a simple calculation based on the calculations of Elci et al.[4]

To study the effect of carrier diffusion alone, we assume that the plasma is in thermal equilibrium with the lattice. This is the case for delay times greater than 100 psec, since on these time scales phonon-assisted relaxation is essentially complete. This plasma , although inhomogeneous in spatial distribution, has a uniform temperature $T = T_\ell$ (lattice temperature). We will assume local charge neutrality and ambipolar diffusion. Thus, electron and hole densities are equal everywhere. Electrons and holes have different Fermi energies E_F and E_H, respectively, which are space and time dependent, although E_F and E_H are related since electron and hole densities are equal. Finally, we will assume that the electron (or hole) density is space and time dependent only via E_F (or E_H). Since diffusion is longitudinal, we have a one-dimensional diffusion problem, where $E_F = E_F(x,t)$ and $E_H = E_H(x,t)$. The electron density and hole density are given by[10]

$$n(x,t) = \int_o^\infty dE \rho_e(E) F\left[\frac{E + E_G - E_F}{k_B T}\right] = \int_o^\infty dE \rho_h(E) F\left[\frac{E + E_H}{k_B T}\right]. \quad (1)$$

Definitions of the various symbols and functions used in this and the subsequent formulae are given in Tables I and II. Also the carrier density, n(x,t), must satisfy the diffusion equation

$$\frac{\partial n}{\partial t} = D \frac{\partial^2 n}{\partial x^2} \quad . \tag{2}$$

We will not attempt to calculate the ambipolar diffusion constant D. Instead, we will use the value measured by Shank and Auston.[8] If an initial density n(x,o) is given, (1) and (2) determine the subsequent $E_F(x,t)$ (or $E_H(x,t)$).

Since the optical pulses used for probing are weak, we can neglect the perturbation of the distributions induced by the probe in the calculation of its transmission. Also, since the probe pulse duration is short compared to the diffusion time scale, we can neglect the effect of the finite probe duration on the probe transmission as well. Within these approximations, the probe transmission is given by

$$T_{pr}(t) \simeq (1 - R)^2 e^{-\int_o^L dx \alpha(x,t)} \quad , \tag{3}$$

where $\alpha(x,t)$ is the absorption coefficient, and $\alpha(x,t)$ depends on x and t via E_F and E_H. The absorption coefficient α as a function of E_F and E_H is calculated in Ref. 4. It is composed of two parts:

$$\alpha(x,t) = \alpha_D(x,t) + \alpha_{FCA}(x,t) \quad , \tag{4}$$

where α_D is due to the direct electronic transitions from the valence bands to the central conduction band valley of Ge and α_{FCA} is due to the free carrier absorption. Expressions for α_D and α_{FCA} are given by Elci et al.[4] as

$$\alpha_D(x,t) = \alpha_o[1 - F(\mu_1) - F(\mu_2)] \tag{5}$$

$$\alpha_{FCA}(x,t) = \alpha_1[\zeta_o(\mu_3) + 2\left(\frac{m_c}{m_h}\right)^2\zeta_o(\mu_4)] + \alpha_2[\zeta_1(\mu_3) + 2\left(\frac{m_c}{m_h}\right)^{5/2}\zeta_1(\mu_4)] \quad . \tag{6}$$

In order to calculate the carrier distribution $n(x,t)$ at some later time, we need the initial distribution $n(x,t_{delay}=0)$, the spatial distribution of carriers as they are generated by the excitation pulse. We are currently working on a model that will give $n(x,t_{delay}=0)$. However, for the present, to avoid considering the temperature dependence of the electron-hole plasma, we take the time origin to be such that temperature relaxation is complete. Thus, $n(x,0) \equiv n(x,t_{delay}\cong 100\text{psec})$ is used as the initial condition for solving the diffusion equation. Of course, $n(x,t_{delay}\cong 100\text{psec})$ is itself not known; we must assume an initial distribution and investigate the effect of its temporal behavior on the sample transmission. Again, we emphasize that we do not intend this to be an exact calculation, but rather a simple investigation of the possible effects of diffusion for long time delays.

To understand the effects of diffusion on the probe transmission, we perform the following calculation. Given the average density of carriers created by the excitation pulse in the interaction region, $\bar{n}$, (see Elci et al.[4]) and assuming a spatially dependent initial carrier density n(x,0), the diffusion equation, Eq. (2), can be solved for the subsequent carrier distribution n(x,t). Results of a typical calculation are shown in Fig. 7., where an initial distribution of the form $n(x,0) = n_o e^{-x/W}$ is assumed. Since the electron temperature is equal to the lattice temperature, the Fermi energies $E_F(x,t)$ and $E_H(x,t)$ can be calculated from Eq. (1). Next, knowing the Fermi energies, α_D and α_{FCA} can be calculated from Eq. (5) and Eq. (6), and the transmission can be calculated from Eq. (3). The spatially dependent direct and free carrier absorption coefficients, calculated from the carrier densities of Fig. 7 , are shown in Fig. 8 and Fig. 9 , respectively. Note that the direct absorption coefficient is always much larger than the free carrier absorption coefficient. Finally, Fig. 10 shows the resulting probe transmission for times greater than 100 psec and for various initial density distributions. The average electron density for all curves in Fig. 7-10 is $2.5 \times 10^{19}/cm^3$.

Figure 10 illustrates that for low average carrier densities, diffusion will tend to decrease the probe transmission, as might be expected. However, when $\bar{n}$ is large enough, diffusion can actually cause an increase in probe transmission, as shown in Fig. 11. While this may seem surprising at first, it becomes more reasonable if the following physical argument is

considered. Consider a simple schematic of the diffusion process such as the one shown in Fig. 12. Since $\bar{n}$ stays constant, the probe pulse sees the same number of carriers independent of time delay. However, the states that are optically coupled are restricted to narrow regions in energy in the band structure. So, if we consider only direct absorption of the probe, not all carriers are effective in filling the optically coupled states and preventing absorption. Thus, it is not right to say that α_D^{-1} for the probe is proportional to $\int_o^L n(x,t)dx$, the total number of carriers, rather,

$$\alpha_D^{-1} \propto \int_o^L n_{eff}(x,t)\, dx = N_{eff} \,, \tag{7}$$

where N_{eff} denotes the total number of carriers effective in clogging the optically coupled states needed for absorption.

N_{eff}, the total number of carriers effective in preventing light absorption, is altered by diffusion as shown schematically in Fig. 12. If the number of carriers is large, when they migrate from the front surface, they can fill the states needed for absorption away from the front surface without depleting the optically coupled states near the surface; N_{eff} increases and absorption decreases. If $\bar{n}$ is small, diffusion will decrease n_{eff} in the front region without increasing n_{eff} significantly in the back, thus the transmission decreases. Thus, depending on the total carrier number, <u>longitudinal</u> diffusion

can cause the probe transmission to rise or fall. Experimentally, we have not observed a rise for long probe delays at present densities.

Diffusion should also affect the probe transmission for shorter delay times as well. That is, carriers diffuse while they lose energy to the lattice. However, when the carriers are hot, D is smaller than its thermal equilibrium value for the same density. In a precise calculation, one would have to allow the diffusion coefficient D to depend on time through the carrier temperature.

Excitation-probe measurements have been performed on $Hg_{1-x}Cd_xTe$[3] as well as Ge. The results of these measurements are presented in Fig. 13 for comparison with the Ge results and for completeness.

IV. SUMMARY

In this paper, we have reviewed recent experiments that measure the nonlinear, nonequilibrium properties of germanium and mercury cadmium telluride. In addition, we have described how the germanium experiments can be accounted for by a recently proposed model that describes the generation and temporal evolution of hot electron-hole plasmas produced in semiconductors by the absorption of intense, ultrashort optical pulses. In these calculations, however, several assumptions, omissions, and simplifications are made. The major assumptions are the following. The carrier distributions are assumed to be Fermi-Dirac for all time scales of interest. All carrier densities and Fermi energies are allowed to depend on time but not position. The overall good agreement between the theory and present experiments indicate that these assumptions work quite well despite the intensities and short durations of the optical pulses. However, more exact calculations and definitive experiments are in progress in an attempt to investigate the implications of these approximations. These calculations will also include processes omitted from the original model, such as Auger recombination and diffusion. We believe it is clear that future investigations with picosecond optical pulses will yield information concerning the individual characteristics of semiconductors and the dynamics of the fundamental processes that take place in semiconductors.

REFERENCES

*Supported by the Office of Naval Research and the North Texas State University Faculty Research Fund.

1. C. J. Kennedy, J. C. Matter, A. L. Smirl, H. Weichel, F. A. Hopf, S. V. Pappu, and M. O. Scully, Phys. Rev. Lett. 32, 419 (1974).
2. A. L. Smirl, J. C. Matter, A. Elci, and M. O. Scully, Opt. Commun. 16, 118 (1976).
3. J. C. Matter, A. L. Smirl, and M. O. Scully, Appl. Phys. Lett. 28, 507 (1976).
4. A. Elci, M. O. Scully, A. L. Smirl, and J. C. Matter, Phys. Rev. B 16, July 1, 1977
5. H. M. van Driel, J. S. Bessey, and R. C. Hanson, to be published in Opt. Commun.
6. H. M. van Driel, A. Elci, J. S. Bessey, and M. O. Scully, Solid State Commun. 20, 837 (1976).
7. W. P. Latham, Jr., A. L. Smirl, A. Elci, and J. S. Bessey, to be published in Solid-State Electron.
8. D. H. Auston and C. V. Shank, Phys. Rev. Lett. 32, 1120 (1974).
9. C. V. Shank and D. H. Auston, Phys. Rev. Lett. 34, 479 (1975).
10. In Eq. (1) and subsequent equations, the x-valleys of Ge are neglected. Since the electron temperature is taken as 300K, their contribution can be neglected without much error.

TABLE CAPTIONS

I. Definition of Symbols

II. Definition of Integrals and Constants

TABLE 1

ω	light frequency	$\hbar\omega$ = 1.17 eV
e	electronic charge	
ϵ_∞	high frequency dielectric constant	16
c	speed of light	
c_Λ	sound velocity	5.4×10^5 cm-sec^{-1}
m	electronic mass	
m_o	electronic effective mass in the central valley of the conduction band	0.1 m
m_c	density of states effective mass in L-valleys	0.22 m
m_h	effective hole mass	0.34 m
T	temperature	300°K
E_o	direct gap	0.889 eV (at 300°K)
E_G	indirect gap	0.664 eV (at 300°K)
Λ_u, Λ_d	electron-acoustic phonon coupling coefficients	17 eV, -3.4 eV
Q_o	electron-optical phonon coupling coefficient	6×10^{-4} erg cm^{-1}
ρ	mass density	5.3 gm-cm^{-3}
Ω_o	optical phonon frequency	$\hbar\Omega_o \simeq 0.02$ eV
R	reflection coefficient	0.34
ℓ	sample thickness	5.5 μm
D	diffusion constant	230 cm^2-sec^{-1}

TABLE II

$$F(z) = \frac{1}{(1 + e^{z})}$$

$$\rho_e(z) = \frac{2^{5/2} m_c^{3/2} z^{1/2}}{\pi^2 \hbar^3}$$

$$\rho_h(z) = \frac{2^{3/2} m_h^{3/2} z^{1/2}}{\pi^2 \hbar^3}$$

$$\zeta_j(z) = \frac{1}{(\hbar\omega)^{3+j/2}} \int_0^{\infty} dy \left[\left(\sqrt{y + \hbar\omega} + \sqrt{y}\right)^{4+j} - \left(\sqrt{y + \hbar\omega} - \sqrt{y}\right)^{4+j} \right] \times \left[F\left(z + \frac{y}{k_B T}\right) - F\left(z + \frac{h\omega + y}{k_B T}\right) \right]$$

$$\mu_1 = \left(\frac{1}{k_B T}\right) \left[E_o + \frac{m_h}{m_o + m_h} (\hbar\omega - E_o) - E_F \right]$$

$$\mu_2 = \left(\frac{1}{k_B T}\right) \left[\frac{m_o}{m_o + m_h} (\hbar\omega - E_o) + E_H \right]$$

$$\mu_3 = \frac{E_H}{k_B T}$$

$$\mu_4 = \frac{E_G - E_F}{k_B T}$$

$$\alpha_o = \frac{2^{5/2} e^2 m_o^{1/2} m_h^{1/2} E_o (h\omega - E_o)^{1/2}}{3\varepsilon_\infty^{1/2} h^3 \omega c (m_o + m_h)^{1/2}}$$

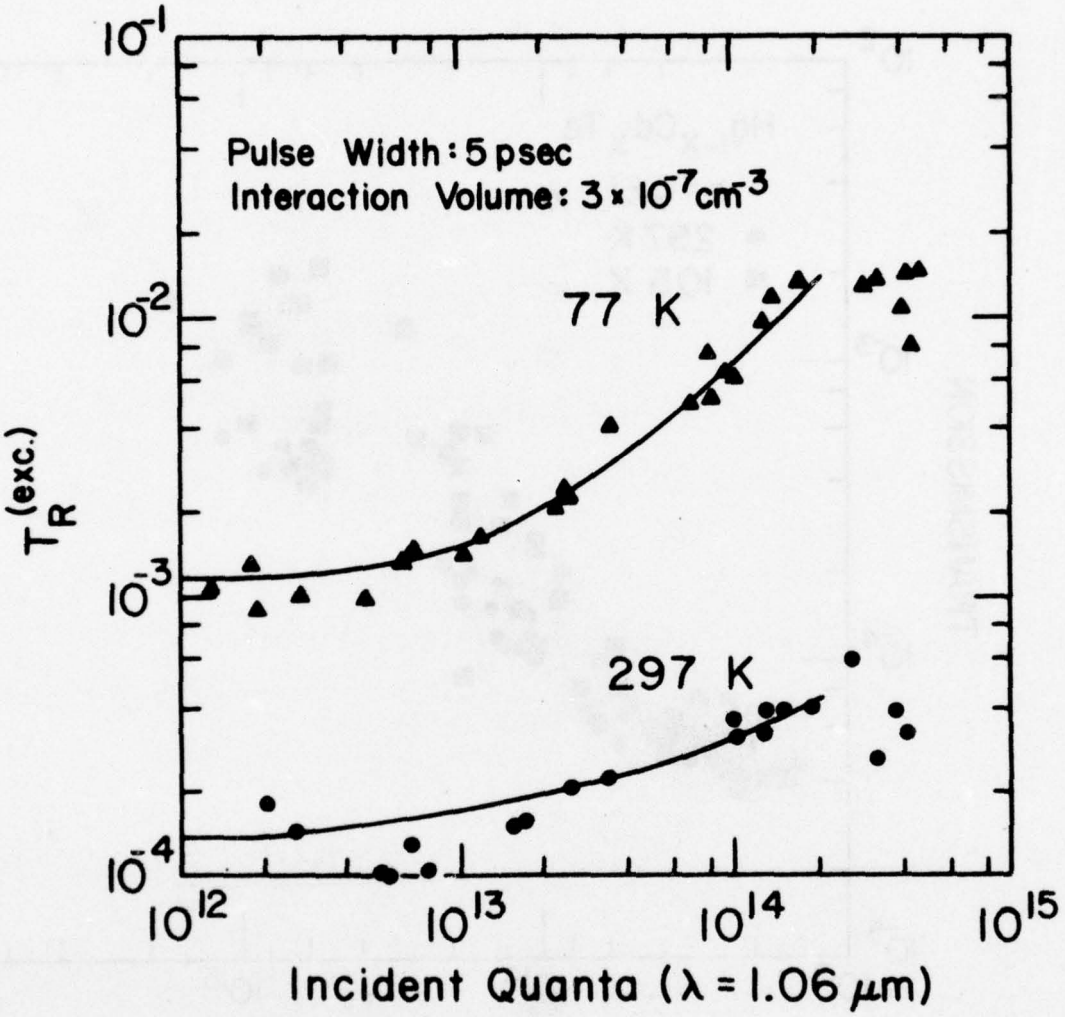
$$\alpha_1 = \frac{e^2 m_h^2}{3\pi^2 \varepsilon_\infty^{1/2} c\rho \hbar^3} \quad \frac{Q_o^2}{\hbar\Omega_o} \left(1 + \frac{2k_B T}{\hbar\Omega_o}\right) + \frac{2k_B T}{3\hbar^2 C_A^2} \left(\Lambda_u^2 + 2\Lambda_u \Lambda_d + 3\Lambda_d^2\right)]$$

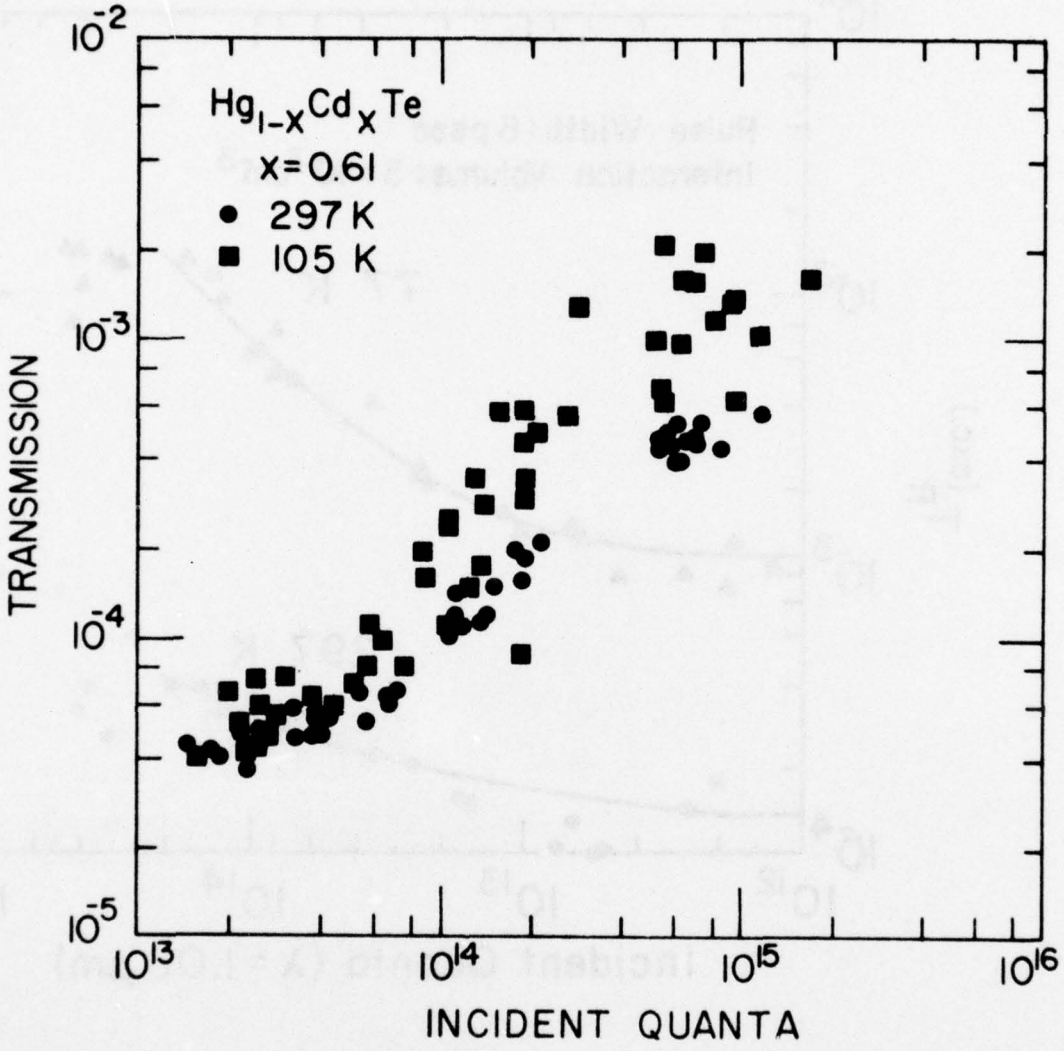
$$\alpha_2 = \frac{2^{5/2} e^2 m_h^{5/2} \omega^{1/2} (\Lambda_u^2 + 2\Lambda_u \Lambda_d + 3\Lambda_d^2)}{45\pi^2 \varepsilon_\infty^{1/2} c\rho c_A \hbar^{9/2}}$$

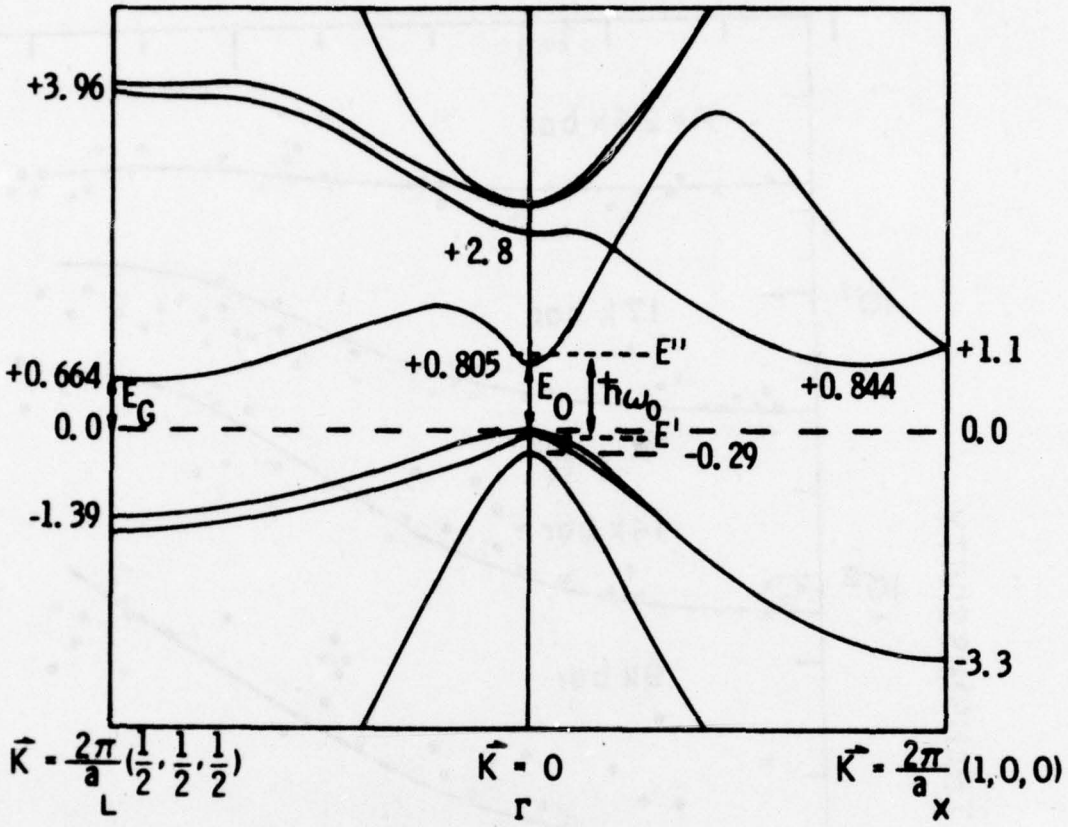
FIGURE CAPTIONS

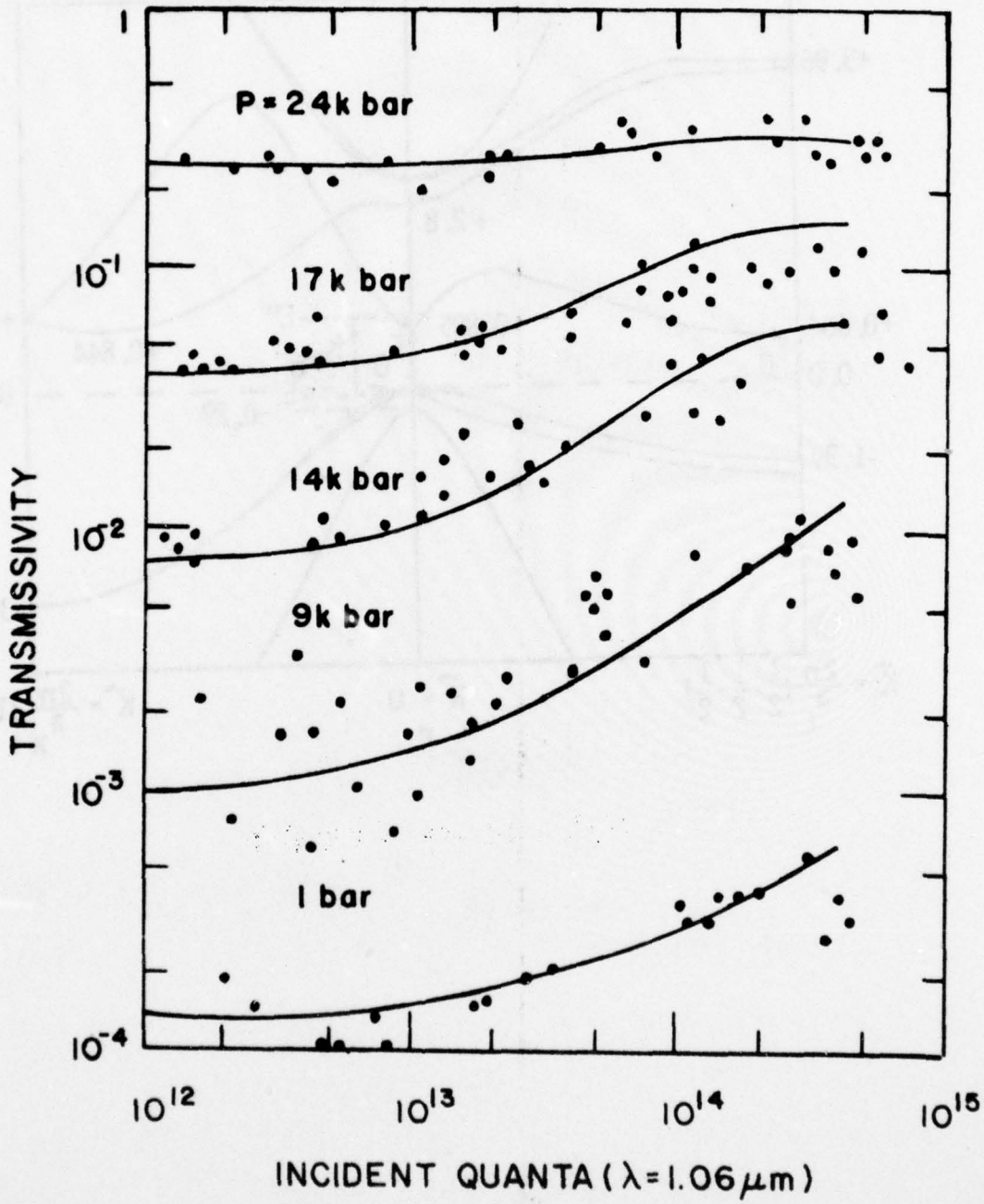
1. Single pulse transmission of a 5.2μm-thick single crystal germanium sample versus incident energy in units of quanta at 1.06μm for temperatures of 297K and 77K.
2. Single pulse transmission of a 5μm-thick $Hg_{0.39}Cd_{0.61}Te$ sample versus incident energy in units of quanta at 1.06μm for temperatures of 297K and 105K.
3. Ge energy-band (eV) structure at 300K.
4. Transmission of a 5.2μm-thick germanium sample as a function of incident energy and pressure, for a sample temperature of 295K.
5. Emission spectra of germanium for different excitation irradiances I_o at temperatures of (i) 295K and (ii) 174K.
6. Probe pulse transmission of a 5.2μm-thick germanium sample versus time delay for sample temperatures of 297K and 77K. Relative transmission units are given by the ratio of the probe transmission to excitation transmission for each delay, normalized such that the peak of the 77K curve is unity.
7. The variation due to diffusion of the carrier density, n(x,t) as a function of longitudinal position and time in a 5.5μm-thick germanium sample. The average density is determined by the total carrier number/interaction volume.
8. The variation due to diffusion of the direct absorption coefficient as a function of longitudinal position and time in a 5.5μm-thick germanium sample.

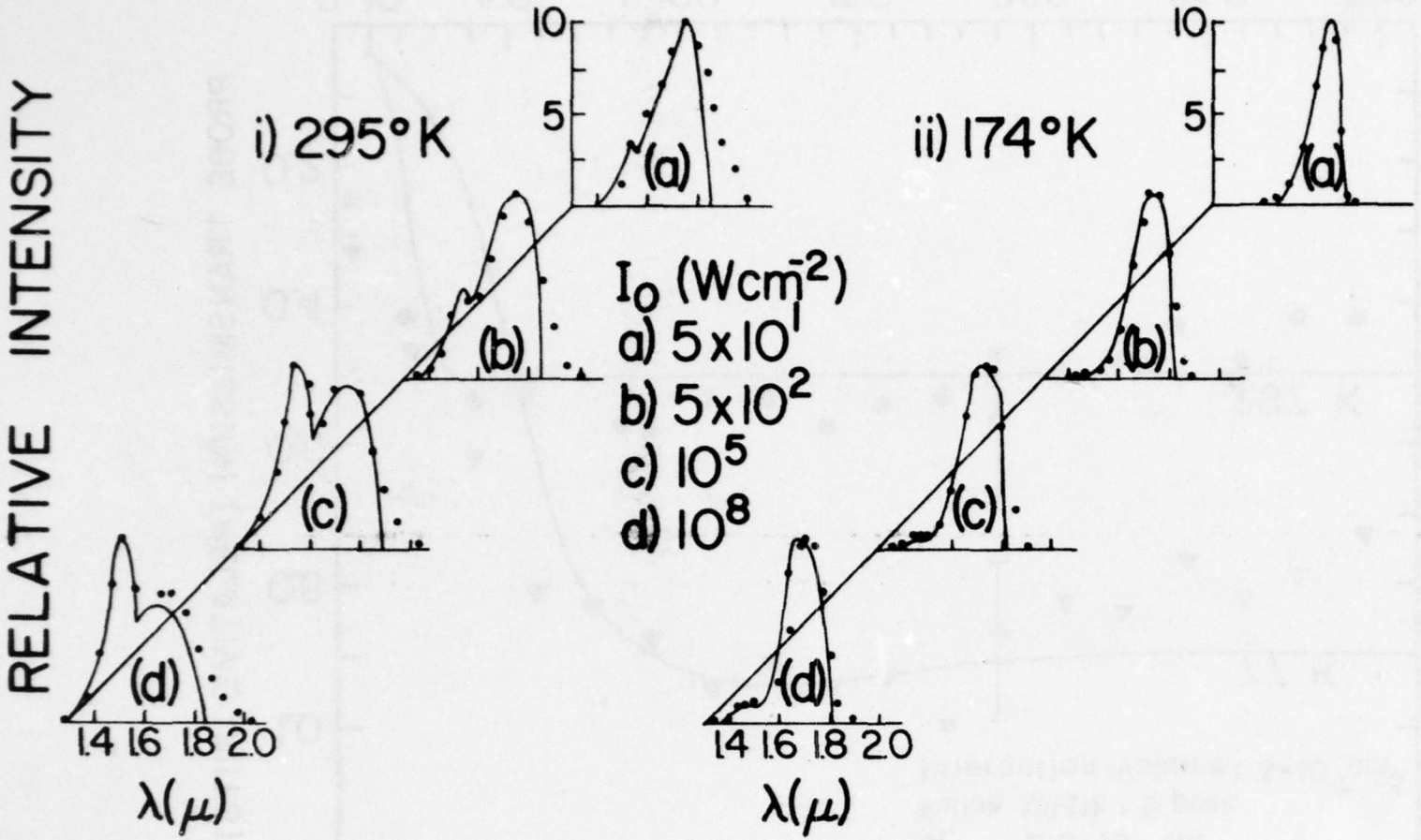
9. The variation due to diffusion of the free carrier absorption coefficient as a function of longitudinal position and time in a 5.5μm-thick germanium sample.
10. The variation of the probe transmission with time as a result of diffusion, for an average carrier denisty of $2.5 \times 10^{19}/cm^3$.
11. The variation of the probe transmission with time as a result of diffusion, for an average carrier density of $4.75 \times 10^{19}/cm^3$.
12. Schematic of the diffusion process
13. Probe pulse transmission of a 5μm-thick $Hg_{0.39}Cd_{0.61}Te$ sample versus time delay for sample temperatures of 297K and 105K.

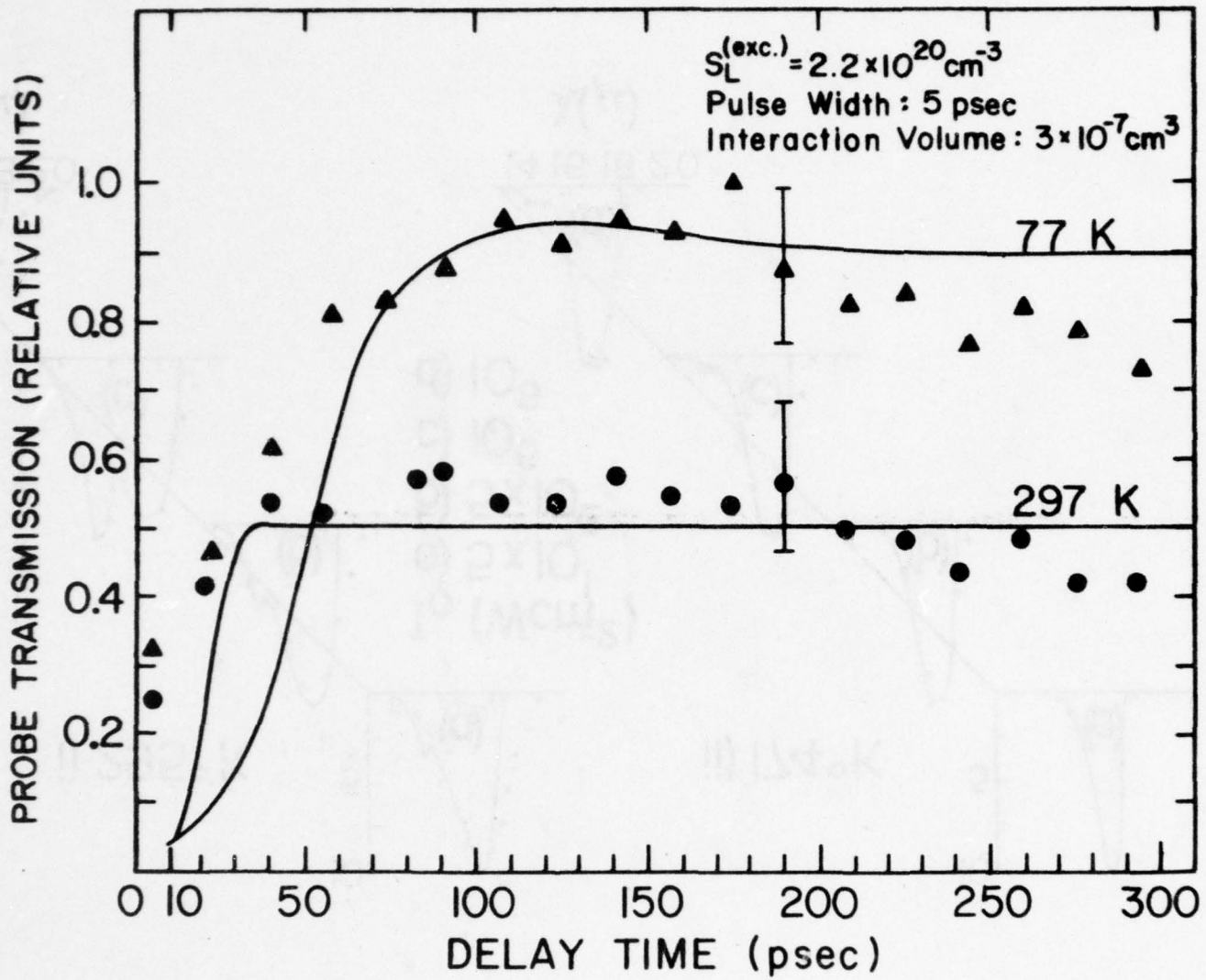

Pulse Width: 5 psec
Interaction Volume: 3 × 10⁻⁷cm⁻³
77 K
297 K
$T_R^{(exc.)}$
10⁻¹
10⁻²
10⁻³
10⁻⁴
10¹²
10¹³
10¹⁴
10¹⁵
Incident Quanta (λ = 1.06 μm)

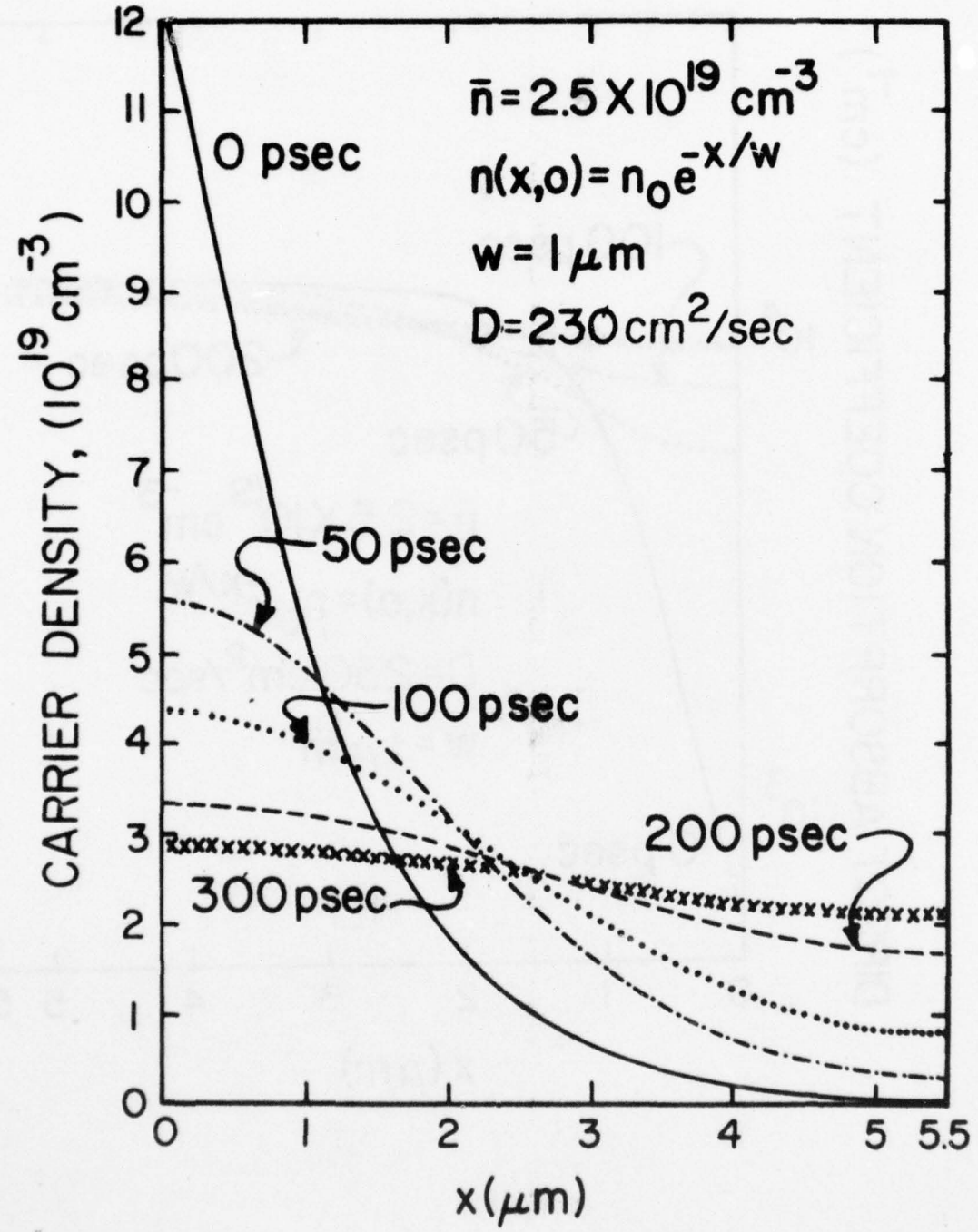

$Hg_{1-x}Cd_xTe$
x=0.61
297 K
105 K
TRANSMISSION
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{13}
10^{14}
10^{15}
10^{16}
INCIDENT QUANTA

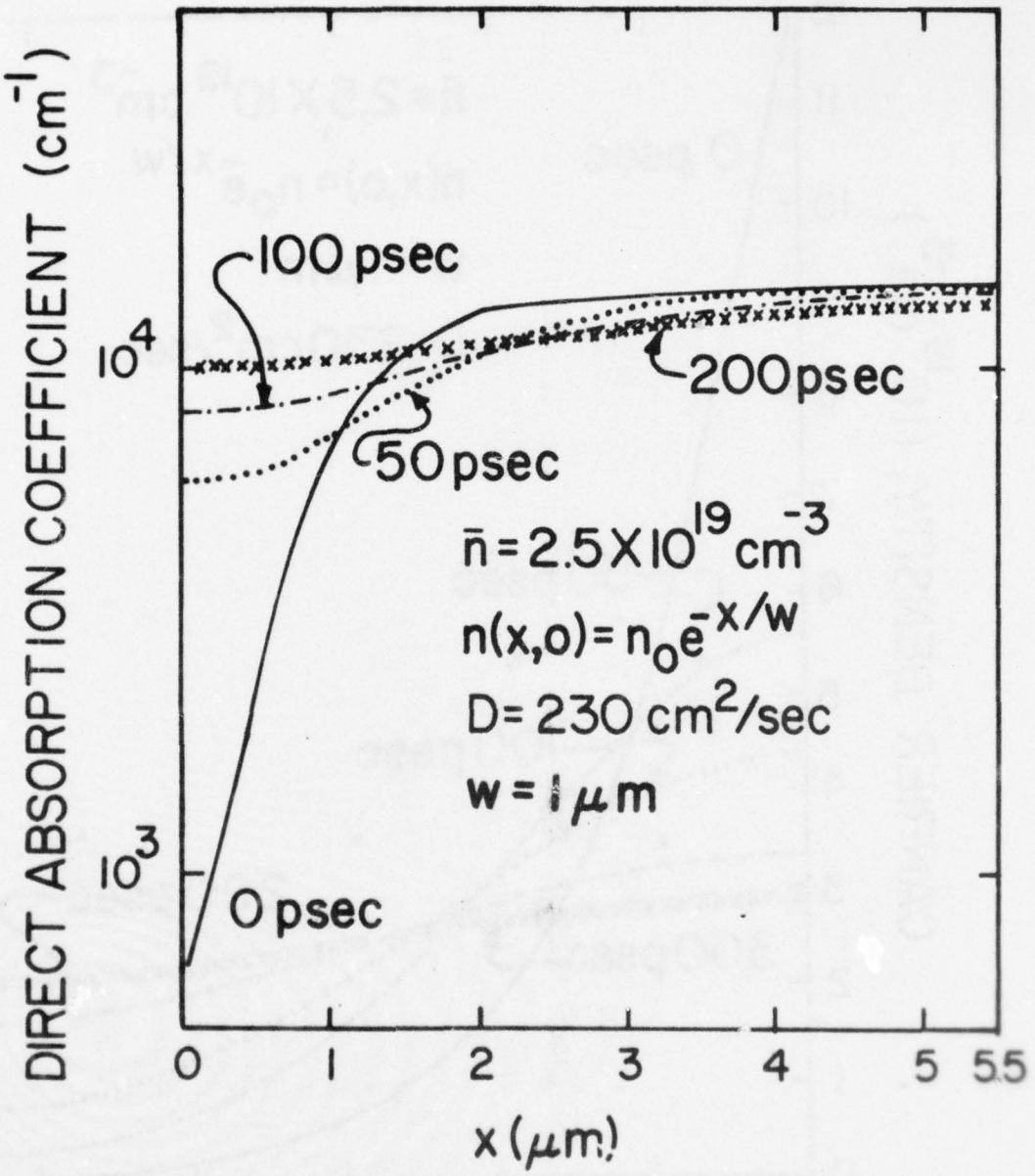

+3.96
+2.8
+0.664
+0.805
E''
+1.1
+0.844
E_G
E_0
ħω_0
0.0
E'
0.0
-0.29
-1.39
-3.3
K = 2π/a (1/2, 1/2, 1/2)
L
K = 0
Γ
K = 2π/a (1,0,0)
X


1
P = 24k bar
10^{-1}
17k bar
14k bar
10^{-2}
9k bar
10^{-3}
1 bar
10^{-4}
TRANSMISSIVITY
10^{12}
10^{13}
10^{14}
10^{15}
INCIDENT QUANTA (λ = 1.06 μm)


RELATIVE INTENSITY
i) 295°K
ii) 174°K
10
5
(a)
(b)
(c)
(d)
I_o (Wcm^{-2})
a) 5×10^1
b) 5×10^2
c) 10^5
d) 10^8
1.4 1.6 1.8 2.0
λ(μ)

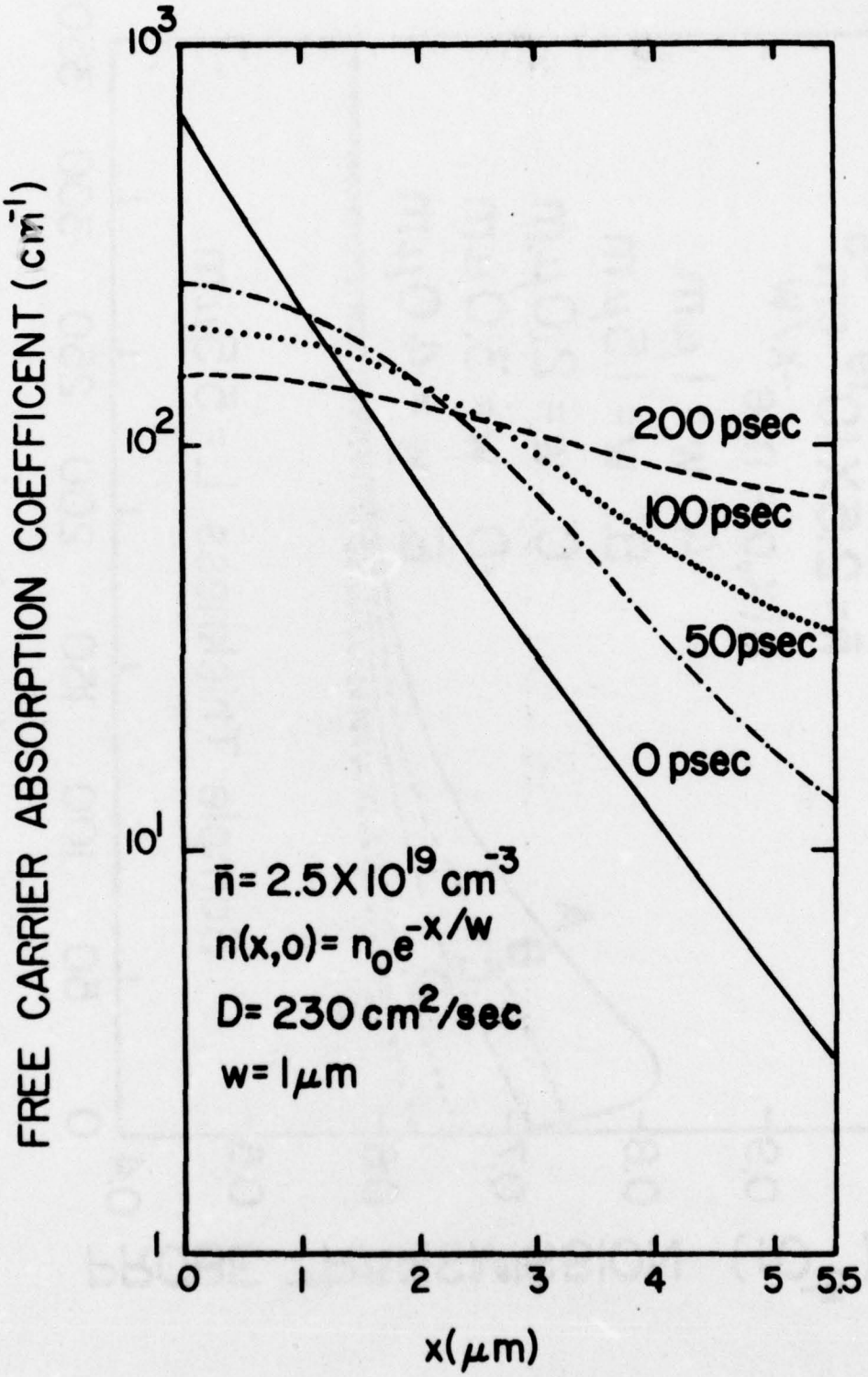

$S_L^{(exc.)} = 2.2 \times 10^{20} cm^{-3}$
Pulse Width : 5 psec
Interaction Volume : $3 \times 10^{-7} cm^3$
77 K
297 K
PROBE TRANSMISSION (RELATIVE UNITS)
1.0
0.8
0.6
0.4
0.2
0 10 50 100 150 200 250 300
DELAY TIME (psec)

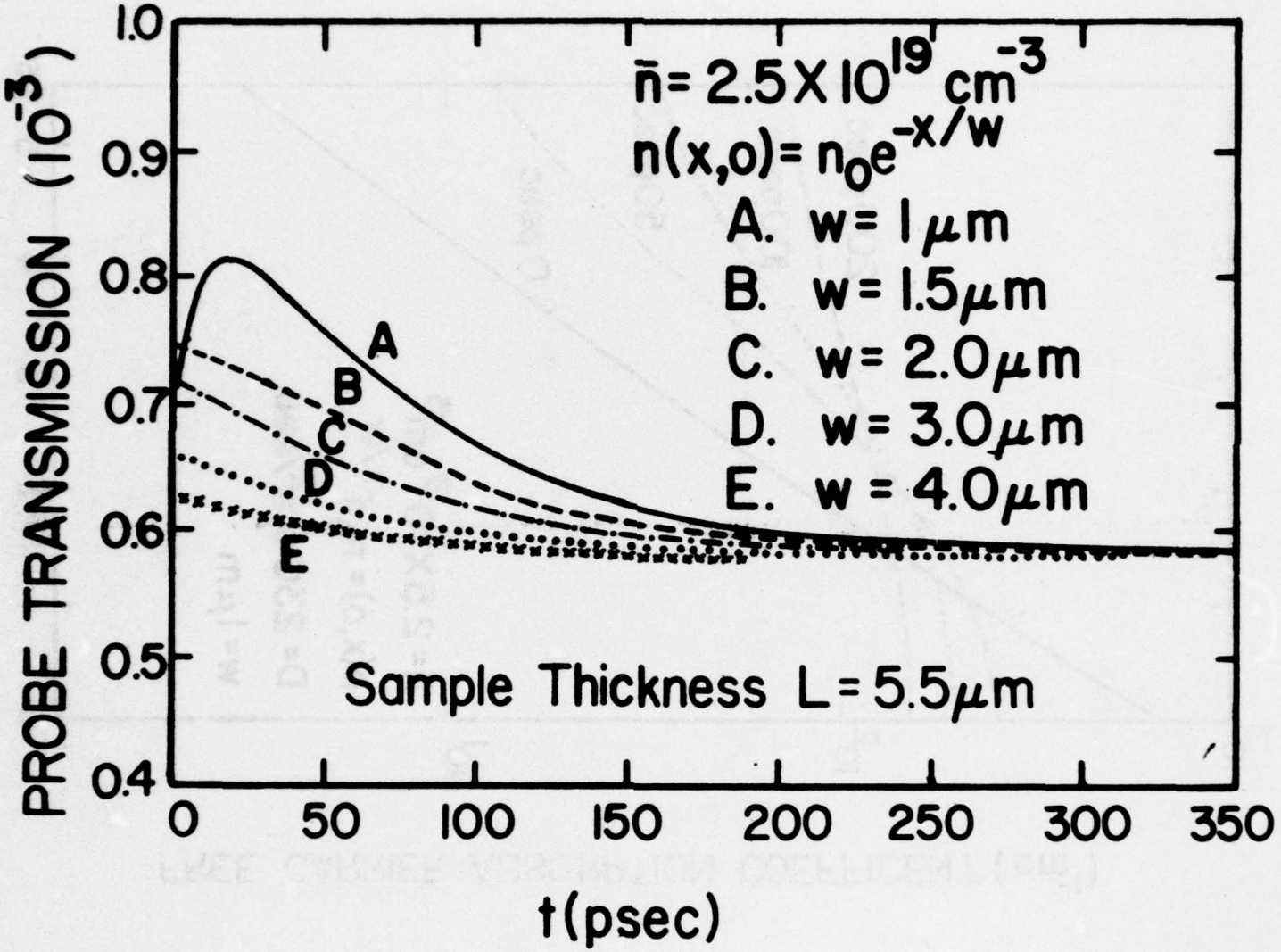

$\bar{n} = 2.5 \times 10^{19}\ cm^{-3}$
$n(x,0) = n_0 e^{-x/w}$
$w = 1\ \mu m$
$D = 230\ cm^2/sec$
0 psec
50 psec
100 psec
200 psec
300 psec
CARRIER DENSITY, $(10^{19}\ cm^{-3})$
$x(\mu m)$
0
1
2
3
4
5
5.5
0
1
2
3
4
5
6
7
8
9
10
11
12

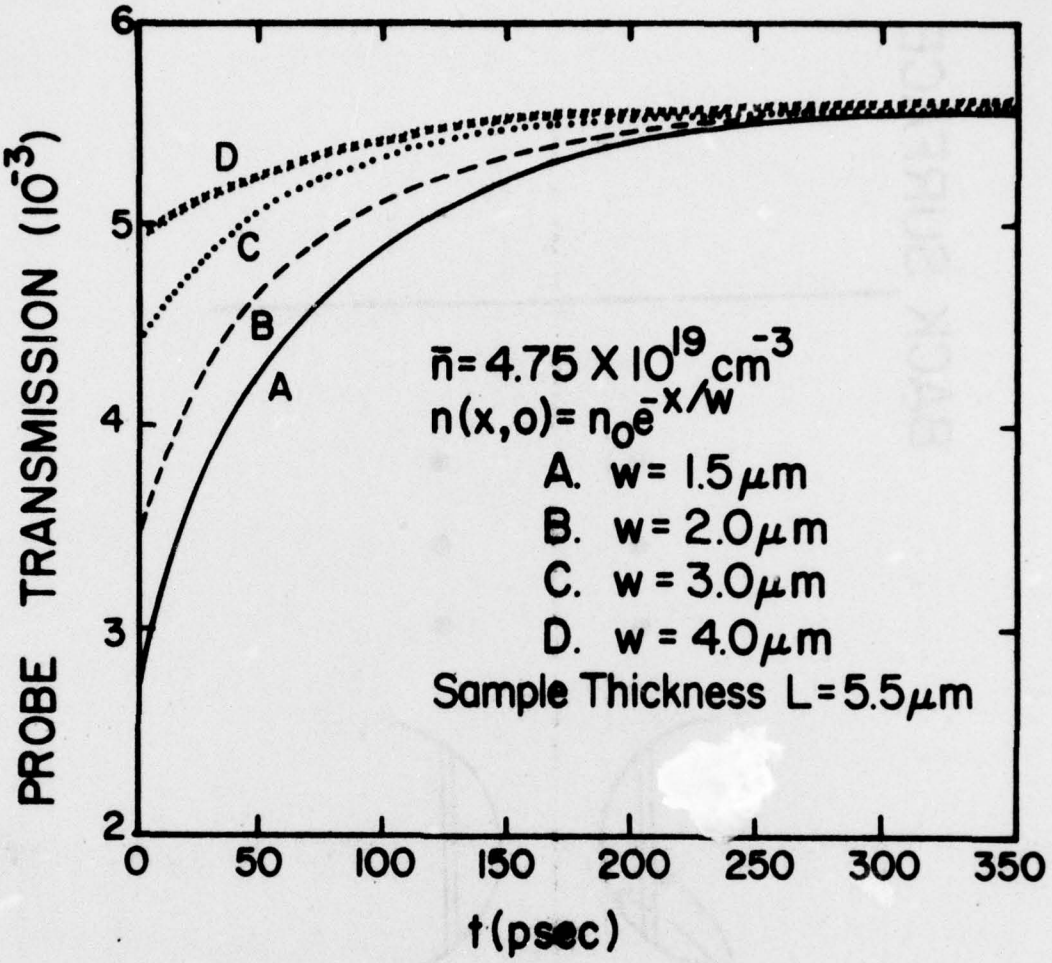

DIRECT ABSORPTION COEFFICIENT (cm⁻¹)
10⁴
10³
100 psec
200 psec
50 psec
0 psec
n̄ = 2.5 X 10¹⁹ cm⁻³
n(x,o) = n₀e^(-x/w)
D = 230 cm²/sec
w = 1 μm
0
1
2
3
4
5
5.5
x (μm)

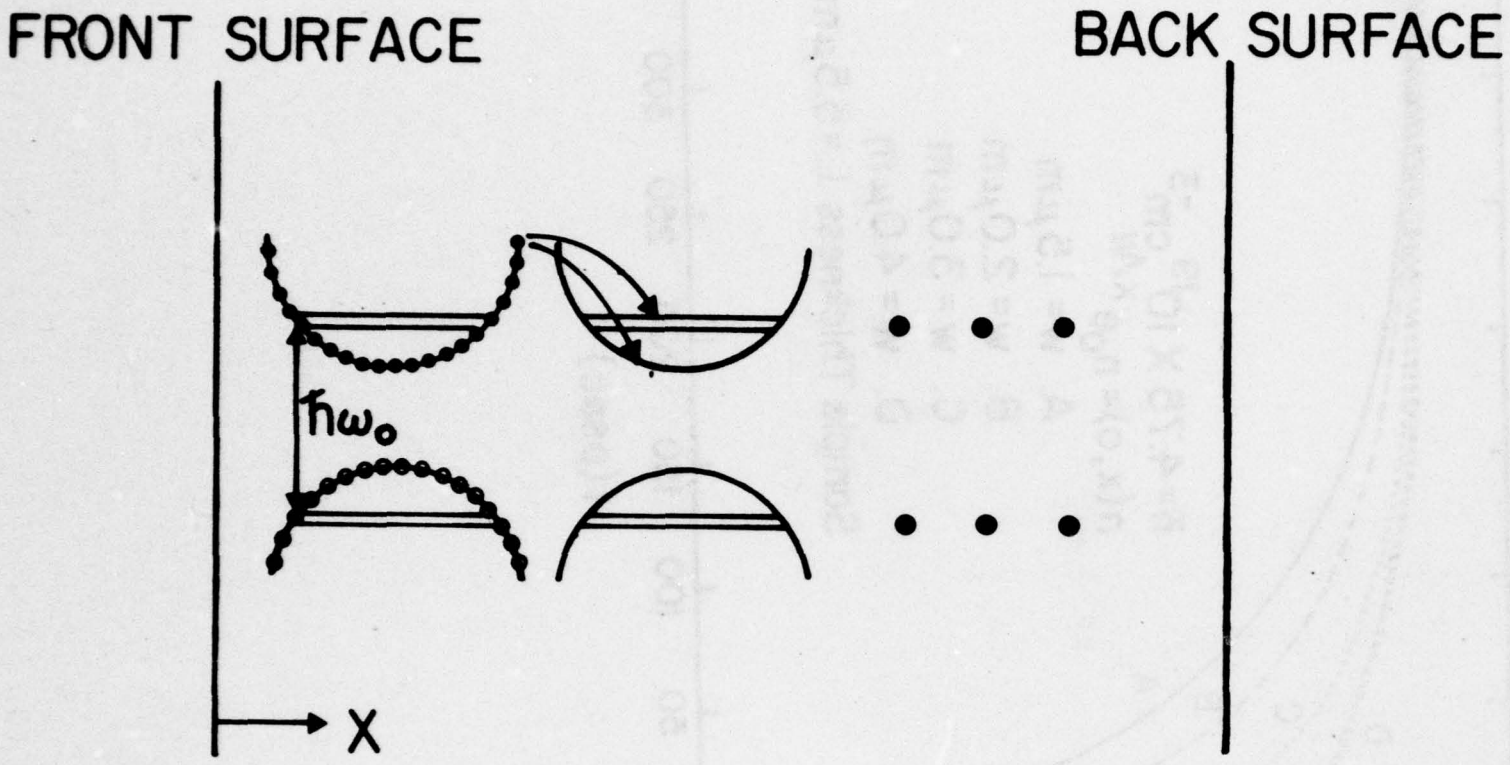
AD-A048 194 NORTH TEXAS STATE UNIV DENTON DEPT OF PHYSICS F/G 20/12
OPTICALLY INDUCED HOT ELECTRON EFFECTS IN SEMICONDUCTORS.(U)
AUG 77 A L SMIRL N00014-76-C-1077
UNCLASSIFIED NL

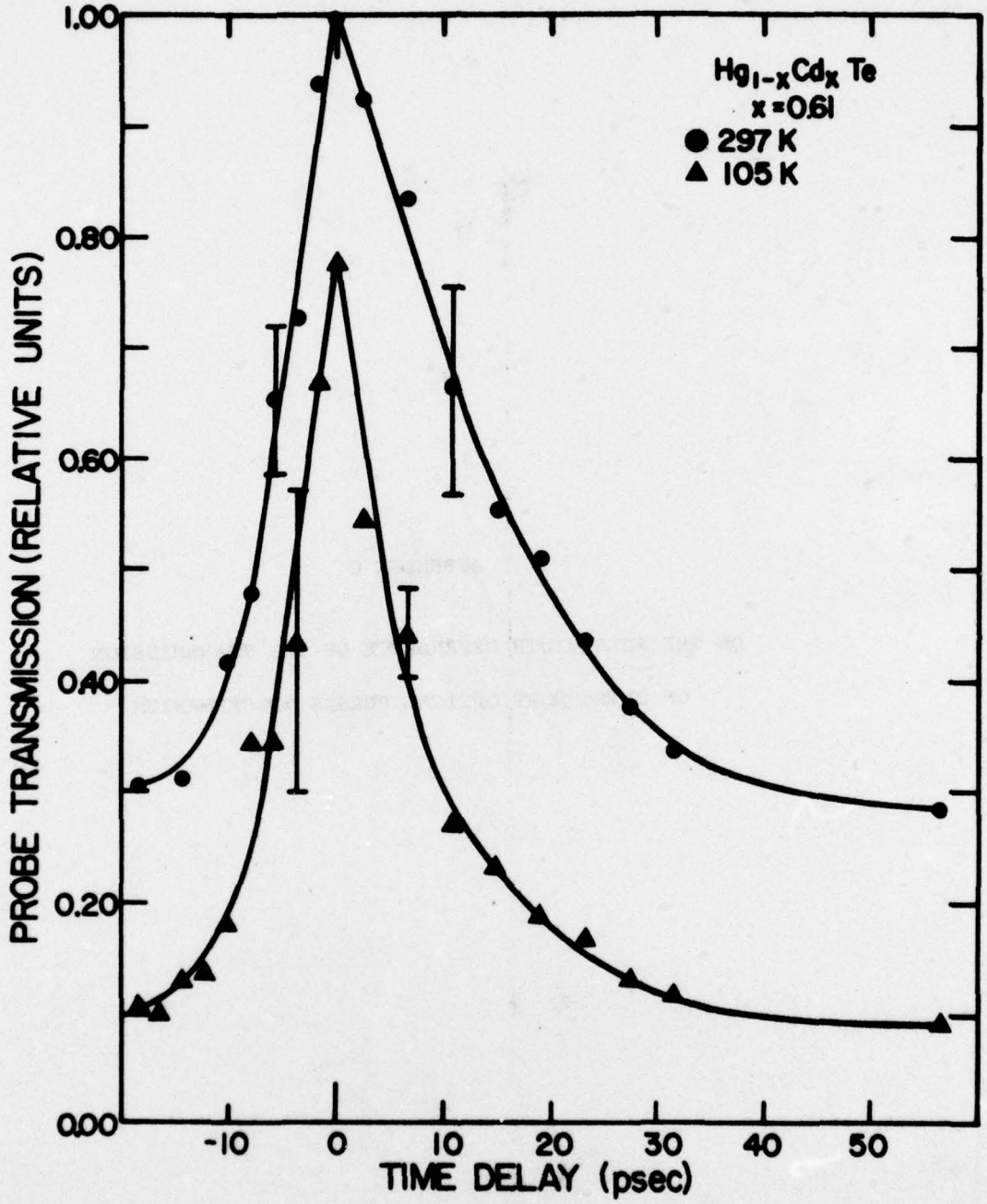
2 OF 3
AD A048194


FREE CARRIER ABSORPTION COEFFICENT (cm⁻¹)
10³
10²
10
1
0
1
2
3
4
5
5.5
x(μm)
200 psec
100 psec
50 psec
0 psec
n̄= 2.5X10¹⁹ cm⁻³
n(x,0)= n₀e^-x/w
D= 230 cm²/sec
w= 1μm


PROBE TRANSMISSION (10^{-3})
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
50
100
150
200
250
300
350
t(psec)
$\bar{n} = 2.5 \times 10^{19}$ cm^{-3}
$n(x,0) = n_0 e^{-x/w}$
A. w = 1 μm
B. w = 1.5 μm
C. w = 2.0 μm
D. w = 3.0 μm
E. w = 4.0 μm
Sample Thickness L = 5.5 μm
A
B
C
D
E


PROBE TRANSMISSION (10^{-3})
6
5
4
3
2
0
50
100
150
200
250
300
350
t(psec)
D
C
B
A
$\bar{n}$= 4.75 X 10^{19} cm^{-3}
$n(x,o)= n_o e^{-x/w}$
A. w= 1.5μm
B. w= 2.0μm
C. w= 3.0μm
D. w= 4.0μm
Sample Thickness L=5.5μm


FRONT SURFACE
BACK SURFACE
$\hbar\omega_o$
X


1.00
0.80
0.60
0.40
0.20
0.00
PROBE TRANSMISSION (RELATIVE UNITS)
-10
0
10
20
30
40
50
TIME DELAY (psec)
Hg1-xCdx Te
x = 0.61
297 K
105 K

APPENDIX C

ON THE PULSEWIDTH DEPENDENCE OF THE TRANSMISSION OF ULTRASHORT OPTICAL PULSES IN GERMANIUM

Submitted for Publication

ON THE PULSEWIDTH DEPENDENCE OF THE TRANSMISSION OF ULTRASHORT OPTICAL PULSES IN GERMANIUM*

John S. Bessey, Bruno Bosacchi and Henry M. van Driel†

Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

and

Arthur L. Smirl

Department of Physics
North Texas State University
Denton, Texas 76203

*Research supported by the Energy Research and Development Administration and the Office of Naval Research

†National Research Council-Canada Postdoctoral Research Fellow. Current address: Erindale College, University of Toronto, Mississauga, Ontario, Canada L5L 1C6

Abstract

The nonlinear transmission of ultrashort optical pulses (λ =1.06μm) through a thin germanium sample has been measured as a function of pulsewidth. These measurements provide insight into a recently proposed theoretical model that describes the enhanced transparency observed at high intensities.

Recently Elci et al.[1] have presented a theoretical model that accounts for the ultrafast transient response of optically-created, hot electron-hole plasmas in germanium. This model is shown to be in substantial agreement with the observed nonlinear transmission of single, ultrashort optical pulses through germanium as a function of incident pulse energy at temperatures of 105 K and 297 K and with the transmission of a weak probe pulse as a function of picosecond time delays after an intense excitation pulse.[2] In addition, the theory agrees well with experimental nonlinear transmission data obtained as a function of band gap energy achieved by pressure tuning a germanium sample.[3] One feature of this model is the predicted strong pulsewidth dependence of the nonlinear transmission of these intense, ultrashort optical pulses through germanium.

This pulsewidth dependence is depicted in Fig. 1. This figure is a plot of the theoretical transmission of a 5.2μm - thick sample of germanium as a function of the number of incident quanta for a sample temperature of 77 K and for theoretical optical pulse widths of 2.5, 5, and 10 psec. Elci et al.[1] have suggested that experiments with variable duration pulses be performed to provide a further test for the theoretical model. In this paper, we report the results of such measurements.

A schematic of the experimental apparatus is shown in Fig. 2. A single pulse, having a wavelength of 1.06μm and a pulse width τ of 5-10 psec, is switched from a train of pulses from a mode-locked Nd:glass laser by an electro-optic shutter. To produce variable-width pulses, this single optical pulse is split into two separate parts that are then recombined collinearly after one part has been delayed with respect to the other with a Michelson interferometer. The recombined pulse is then focused on the sample to a spot with diameter of approximately 250μm. The sample

is an undoped, high purity (ρ_{min} = 40.0 Ωcm), single crystal germanium wafer that is cut with the (111) plane as the face, mounted on a KZF-2 substrate, and polished to a thickness of 7.5μm, as determined from an interferometric technique.

Figure 3 is a plot of the transmission versus incident pulse energy for two pulse widths, τ and 2τ, and for a sample temperature of 105 K. Each point shown is the average of 8 to 12 data points, and the error bars indicate one standard deviation. Although the exact pulse width τ is uncertain from one optical pulse to the next, it can be readily seen that the difference between the two curves is less than the uncertainties in the data and substantially less than that predicted by theory.

The pulsewidth dependence of the transmission can be understood in terms of this model in the following way. Initially, when the optical pulse enters the Ge sample most of it is absorbed creating a large number of electrons (holes) in the central valley of the conduction (valence) band. These electrons are rapidly (compared to a psec) scattered to the conduction band side valleys by phonons. Carrier-carrier scattering events ensure that the carrier distributions are Fermi-like. Since the energy of the light quanta $\hbar\omega_o$ is greater than the direct band gap E_o and the indirect band gap E_G, a direct absorption event, such as the one described above, gives an excess energy of approximately $\hbar\omega_o - E_G$ to electronic thermal agitation. Thus, direct absorption followed by intervalley phonon and carrier-carrier scatterings causes the conduction (valence) band to contain a large number of electrons (holes) with a temperature of approximately 1500 K.

If only these processes are considered, the Ge transmission will rise as the optical pulse energy is increased because of the saturation of the

available optically coupled states due to direct absorption; but no dependence on pulsewidth will result. However, intravalley electron-phonon relaxation and plasmon-assisted electron-hole recombination, when included, will cool or heat, respectively, the carrier distribution. If the phonon-assisted cooling dominates, more of the electrons (holes) will be located lower (higher) in the conduction (valence) band, thus more completely filling the states needed for absorption; the transmission will further increase. However, if heating due to plasmon-assisted recombination is the stronger, the electrons (holes) will be located higher (lower) in the conduction (valence) band, thus freeing the states needed for optical absorption; the transmission will increase more slowly as a function optical pulse energy.

To understand the pulsewidth dependence of these two processes, consider two pulses of equal energy but different width incident on a Ge sample. While the longer pulse interacts with the sample, there is more time for both phonon-assisted relaxation and plasmon-assisted recombination. If heating due to plasmon emission is dominant, the longer pulse will have the higher distribution temperature and, consequently, the lower transmission. If the phonon-assisted relaxation is stronger, similar arguments show the transmission will have the opposite pulsewidth dependence. Although we have simplified the situation somewhat, these are the dominant effects. The coupling constants in the paper by Elci et al.[1] were such that phonon-assisted cooling of the carrier distribution was small during the time the pulse was present in the sample, thus plasmon heating effects determined the pulsewidth dependence.

Thus, the present data represents the first substantial disagreement between experiment and a heretofore successful model. (From our discussion

of the pulsewidth dependence of the model it appears that for some reason either the plasmon-assisted recombination is not as strong as we originally calculated or that the phonon-assisted relaxation is stronger). Whether this disagreement is due to the simplifying approximations made in the model, to the experimental uncertainties of physical constants used, or to the details of the physical processes considered in the model is currently being investigated. Because the calculations of Elci et al.[1] have provided a theoretical description of a substantial body of experimental data, we still believe this model to be substantially correct. However, the present measurements indicate the need for further experimental and theoretical study in order to place our understanding of these ultrafast processes on a firm basis.

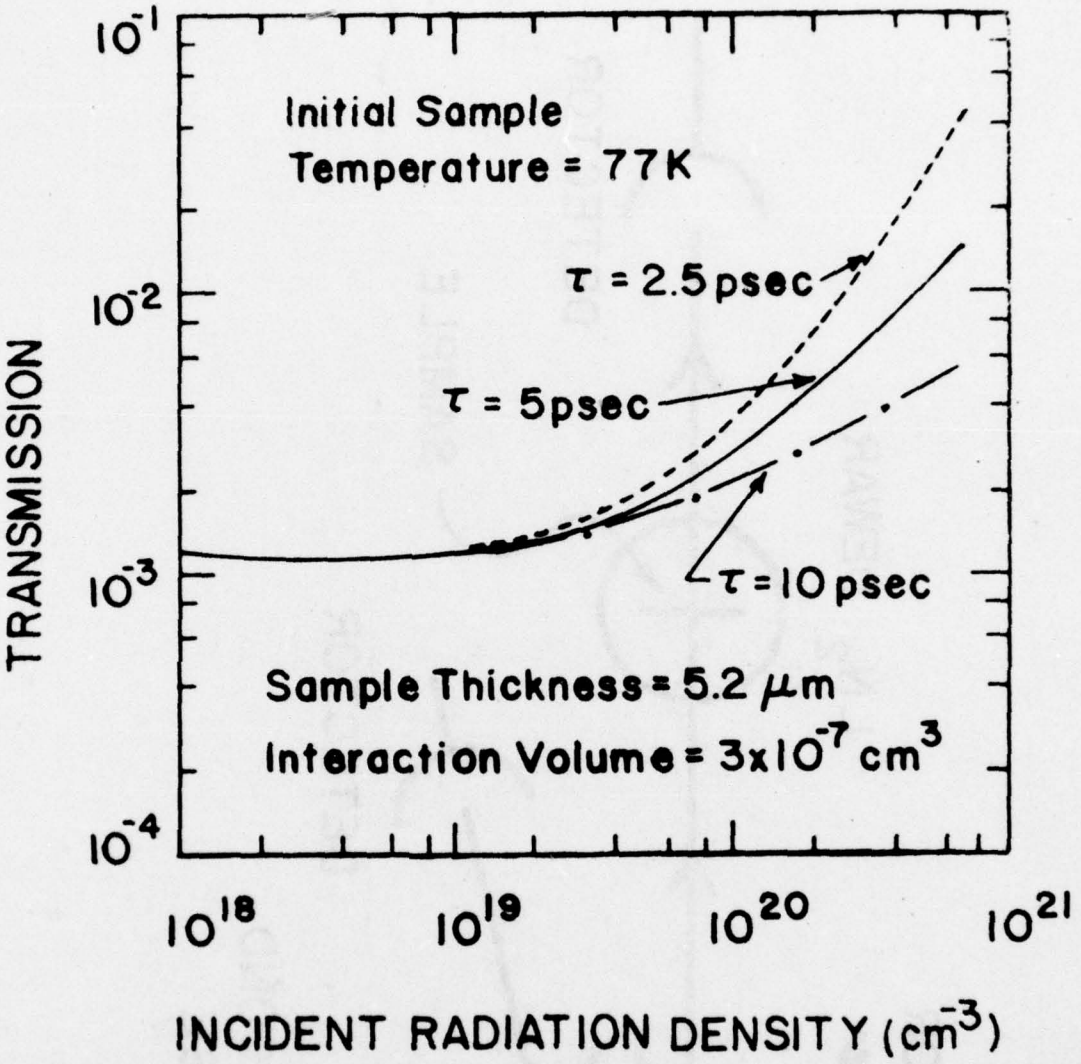
We thank W. P. Latham and J. C. Matter for their comments concerning the manuscript.

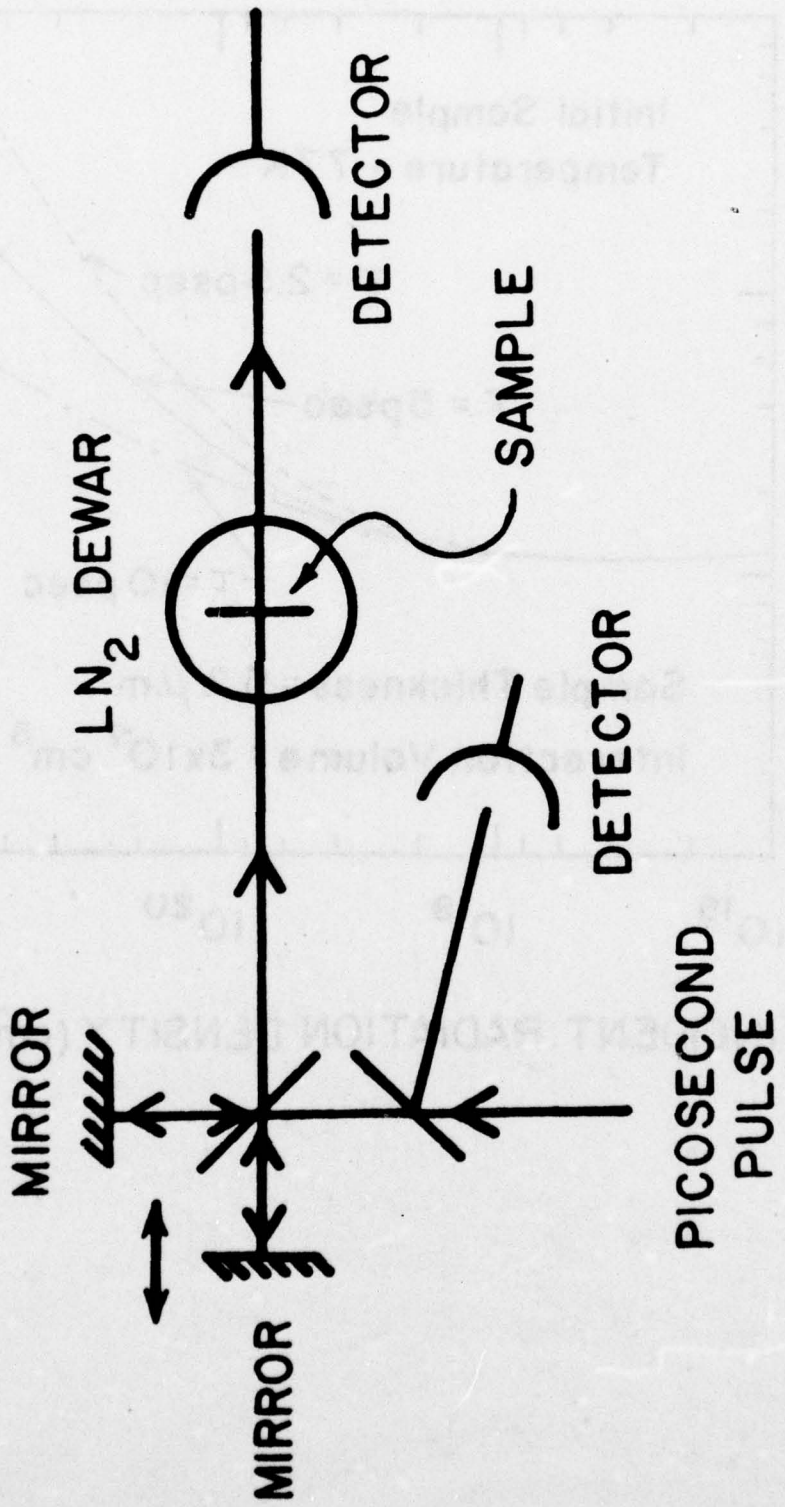
References

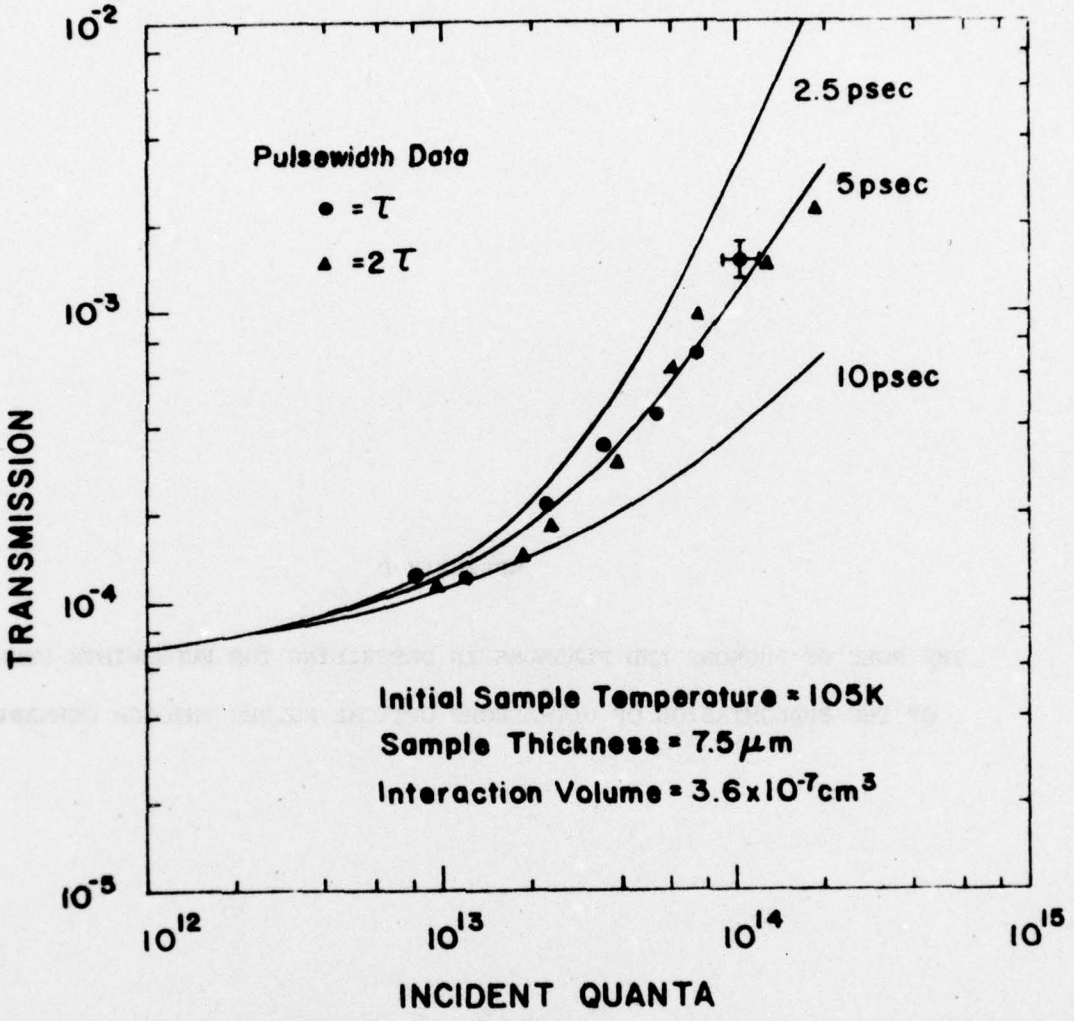
1. A. Elci, M. O. Scully, A. L. Smirl and J. C. Matter, (to be published in July 1, 1977 issue of Phys. Rev. B.)
2. A. L. Smirl, J. C. Matter, A. Elci, and M. O. Scully, Opt. Commun. 16, 118 (1976).
3. H. M. van Driel, J. S. Bessey, and R. C. Hanson, (to be published).

Figure Captions

1. Predicted single pulse transmission of a thin Ge sample vs. incident pulse energy density (number of quanta/interaction volume) for pulsewidths of 2.5, 5, and 10 psec for an initial lattice temperature of 77 K (from Elci et al.[1]).
2. Block diagram of experimental apparatus for producing pulses of variable width.
3. Experimental single pulse transmission of a 7.5μm - thick Ge sample vs. incident pulse energy for pulsewidths of τ and 2τ for a lattice temperature of 105 K. Corresponding theoretical curves are shown for pulsewidths of 2.5, 5, and 10 psec.


Initial Sample
Temperature = 77K
τ = 2.5psec
τ = 5psec
τ = 10psec
Sample Thickness = 5.2 μm
Interaction Volume = 3x10⁻⁷ cm³
TRANSMISSION
10⁻¹
10⁻²
10⁻³
10⁻⁴
10¹⁸
10¹⁹
10²⁰
10²¹
INCIDENT RADIATION DENSITY (cm⁻³)


MIRROR
MIRROR
LN_2 DEWAR
DETECTOR
SAMPLE
DETECTOR
PICOSECOND
PULSE


TRANSMISSION
INCIDENT QUANTA
Pulsewidth Data
● = τ
▲ = 2τ
2.5 psec
5 psec
10 psec
Initial Sample Temperature = 105K
Sample Thickness = 7.5 μm
Interaction Volume = 3.6x10⁻⁷cm³

APPENDIX D

THE ROLE OF PHONONS AND PLASMONS IN DESCRIBING THE PULSEWIDTH DEPENDENCE OF THE TRANSMISSION OF ULTRASHORT OPTICAL PULSES THROUGH GERMANIUM

Accepted for Publication by
Solid State Electronics

THE ROLE OF PHONONS AND PLASMONS IN DESCRIBING THE PULSEWIDTH DEPENDENCE OF THE TRANSMISSION OF ULTRASHORT OPTICAL PULSES THROUGH GERMANIUM*

William P. Latham, Jr., Arthur L. Smirl, and Ahmet Elci
Department of Physics
North Texas State University
Denton, Texas 76203

and

J.S. Bessey
Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

*This work supported by Office of Naval Research

ABSTRACT

Recently, Elci, et al. developed a first principles model that accounts for the generation and transient behavior of dense electron-hole plasmas produced in germanium by picosecond optical pulses. The agreement between this model and early experiments is substantial. However, recent transmission measurements, involving optical pulses of varying width, are in disagreement with initial predictions of this model. Here, we emphasize the dependence of the optical properties of the Ge plasma on the electron-phonon coupling constant, the broadening in the plasmon-assisted recombination, and the energy band structure, and we suggest adjustments to the original model that should produce agreement with these more recent experiments. The pulsewidth dependence in the transmission is shown to be sensitive to the relative strengths of the electron-phonon relaxation and the plasmon recombination during the temporal evolution of the hot electron-hole plasma.

I. INTRODUCTION

Recently, Elci, et al.[1] have proposed a theoretical model that describes the generation and the temporal evolution of hot electron-hole plasmas produced in the elemental semiconductor germanium by the absorption of intense, ultrashort optical (λ = 1.06μm) pulses. This model has been shown to be in agreement with the experimentally-observed enhanced transmission of single, energetic pulses through thin germanium samples as a function of incident pulse energy for sample temperatures of 105 K and 300 K.[2,3] It has also successfully described the transmission of a weak probe pulse through the germanium as a function of time delay after an energetic excitation pulse for different excitation energies and for sample temperatures of 105 K and 300 K.[3] Subsequently, van Driel, et al.[4] have conducted further transmission studies, in which the energy band gap of the germanium sample is tuned by hydrostatic pressure, that corroborate the proposed model.

However, Bessey, et al.[5] have recently experimentally measured the optical pulsewidth dependence of the transmission of thin germanium samples, as suggested by Elci, et al.[6], and have found these measurements to be in substantial disagreement with the predictions of the model. In addition, Elci, et al. have reported[7] that, although normalized, relative transmissions are in agreement with experiment, absolute calculated transmissions are somewhat higher than those

measured. The question thus arises as to whether these disagreements with the proposed model are due to the assumptions made in deriving the model, the experimental uncertainties in physical constants used in the calculations, or the physical processes included in the model.

We demonstrate in this paper that variations produced in the model by attempts to simplify the germanium energy band structure and by the experimental uncertainties in the electron-phonon coupling constants and the strength of the plasmon-assisted electron-hole recombination are sufficient to allow this disagreement between model and experiment.

In Section II, we begin with a brief review of the theoretical model of Elci, et al.[1] Then, in Section III, we demonstrate that, in this model, the absolute optical transmission is drastically affected by replacing the non-parabolic energy band structure of germanium by a parabolic band structure. In Section IV, we show that the optical pulsewidth dependence of the phonon-assisted carrier relaxation processes and the plasmon-assisted recombination processes is opposite and that the pulsewidth dependence of the transmission can be varied at will by adjusting the relative strengths of the two processes.

II. REVIEW OF THE MODEL

According to the model proposed in Ref. 1, the transmission of a single optical pulse through a thin germanium sample as a function of incident pulse energy can be accounted for as follows. When an excitation pulse is incident on a Ge sample, the unreflected portion of the pulse enters the sample where most of it is absorbed by direct transitions, creating a large number of electrons (holes) in the central valley of the conduction (valence) band. The electrons are rapidly ($\tilde{<}$ 10^{-14}sec) scattered to the conduction band side valleys by long wave vector phonons. Particle-particle scattering events, which occur at a rate faster than or comparable to the direct absorption rate, ensure that the carrier distributions are Fermi-like and that both electron and hole distributions have the same temperature, which can be different from the lattice temperature. Since the photon energy $\hbar\omega_o$ is greater than either the direct energy gap E_o or the indirect band gap E_G, such a direct absorption event followed by phonon-assisted scattering of an electron to the side valleys results in the photon giving an excess energy of $\hbar\omega_o - E_G$ to thermal agitation. This excess energy results in an initial distribution temperature (approximately 1500 K for a lattice temperature of 300 K) due to direct absorption that is greater than the lattice temperature. If the above processes

were the only ones considered, the single pulse optical transmission would begin at its Beer's law value and increase as a function of incident pulse energy because of the saturation of the optically coupled states as a result of direct absorption. Note that, owing to the rapid scattering of electrons from the central to side valleys of the conduction band by phonons, any decrease in the number of states available in the central valley for direct absorption is ultimately determined by the buildup of the electron population in the X and L-side valleys.

As the carrier density increases because of direct absorption, the plasma frequency of these carriers increases. At sufficiently large plasma frequencies, an electron in the Γ-valley can recombine with a hole near the top of the valence band via emission of a plasmon. Normally, plasmon-assisted recombination can occur only if the plasma frequency ω_p is larger than the direct band gap frequency $E_o/\hbar$. However, in our case, the plasma resonance is considerably broadened due to the nature of the direct absorption and subsequent scattering between the central and the side conduction band valleys. Therefore, plasmon-assisted recombination can occur at plasma frequencies lower than $E_o/\hbar$. These collective plasma oscillations have lifetimes that are short compared to a picosecond. The energy lost in the decay of the plasmons is rapidly transferred to single electron and hole

states and, ultimately, increases the temperature of the carrier distribution. Therefore, as the electron and hole populations increase, the plasmon-assisted recombination rate increases, which limits the carrier number and raises the electron-hole temperature. Thus, further increase in the sample transmission is prevented as the electrons (holes) are heated and removed from the optically coupled states. Free carrier absorption events, although less important, also serve to increase the distribution temperature while the pulse is incident on the sample.

Owing to direct absorption and plasmon-assisted recombination, the interaction region of the sample contains a large number of carriers with a high electron-hole temperature. These electrons (holes) located high in a conduction (valence) band valley can relax by emitting phonons. The effect of this relaxation is to reduce the electron-hole temperature and increase the lattice temperature. As the distribution cools by means of these intraband phonon-assisted transitions and carriers fill the states needed for absorption, the transmission increases. The phonon-assisted relaxation occurs on a time scale comparable to the pulse width.

After the passage of the optical pulse, the interaction region of the sample contains a large number of carriers with a high distribution temperature. The final temperature is determined by the number of quanta in the optical pulse and

by the relative strengths of the plasmon-assisted recombination and the phonon-assisted relaxation. The plasmon-assisted recombination is essentially turned off when the excitation pulse has passed through the sample. As time progresses, the distribution will continue to cool by phonon-assisted intravalley relaxation. Experimentally, a weak probe pulse interrogates the evolution of the distribution after the passage of the first pulse and is a sensitive measure of whether the optically coupled states are available for absorption or are occupied. Immediately after the passage of the excitation pulse the probe transmission is small since the electrons (holes) are located high (low) in the bands because of the high distribution temperature, leaving the states that are optically coupled available for direct absorption. Later, as the distribution temperature cools and carriers fill the states needed for absorption, the transmission increases.

The position of the optically coupled states needed for direct absorption depends on the model band structure. A discussion of the effects of the simplified band structure on the germanium transmission is given in the next section.

III. SIMPLIFICATIONS TO THE BAND STRUCTURE

One of the basic simplifications in the theoretical calculations of Ref. 1 is to replace the actual germanium energy band structure with a parabolic band structure having two degenerate valence bands and a conduction band with a direct valley having energy gap E_o and ten equivalent side valleys with averaged energy gap E_G. Actually, the central valley in germanium is highly non-parabolic, and the six side valleys in the (1,0,0) direction are located higher in energy than the four side valleys in the (1,1,1) direction. In this section, we investigate numerically the effect of assuming a parabolic central conduction band valley on the absolute value of the optical transmission. We illustrate that such assumptions are sufficient to account for the discrepancy between the calculated and observed transmissions reported in Ref. 1.[7]

If the calculated transmission is to fit accurately the experimental data, the number of available states in the conduction and valence bands must be properly estimated. However, an enormous simplication in the calculations results if the bands are taken to be parabolic. At the band edge, the central valley of the conduction band can be taken to be parabolic with an effective mass $m_o = 0.04m_e$, where m_e is the electron rest mass. However, the photon energy $\hbar\omega_o$ is greater than the direct band gap, and the optically coupled

states are located, in general, above the central valley edge, where the central valley is broader than this effective mass would predict. Given that one wishes to make a parabolic approximation to the band structure, the question then arises as to the "best" choice for the central valley effective mass.

One solution, which has already been mentioned, is to choose the effective mass equal to the band edge effective mass, $m_o = 0.04m_e$. However, this procedure results in an erroneous value for the linear absorption coefficient since it does not accurately reflect the curvature of the bands near the optically coupled states. It also results in an erroneous estimate of the states available up to the optically coupled states.

Another solution is to equate the expression[8] for the Beer's law absorption coefficient calculated assuming parabolic bands to the experimentally measured coefficient at the frequency ω_o and to solve for m_o. This procedure ensures an accurate linear absorption coefficient but still results in an error in estimating the number of states available up to, and including, the optically coupled states. This procedure produces a value of $m_o = 0.06m_e$ at 77 K and $m_o = 0.12m_e$ at 300 K. A higher value of m_o is obtained for the higher temperature, since the band gap is less and the optically

coupled states are located higher in the actual band structure.

Figure 1 and Table I illustrate the magnitude of the error that can be introduced by making a parabolic approximation to the direct conduction band valley. Figure 1 shows the simplified germanium band structure used in the calculations of Ref. 1 for the two effective masses, $m_o = 0.04m_e$ and $m_o = 0.1m_e$, at a temperature of 300 K. The optically coupled states are indicated in the figure. Table I illustrates the effect that these two choices of effective mass can have on the number of available states in the conduction band central and side valleys up to the optically coupled states. Note that, as the central valley effective mass is varied from $0.04m_e$ to $0.10m_e$, the number of electrons required to fill the conduction band up to and including the optically coupled states is not appreciably changed; however, the number of holes required to fill the valence band is tripled.

The choice of the central valley effective mass can have a dramatic effect on the germanium transmission. Figure 2 is a plot of the single pulse transmission of a 5.2μm-thick germanium sample as a function of the incident number of quanta for four effective masses. The sample temperature is 300 K. The curves are generated by using the model presented in Ref. 1 and varying the central valley effective mass, m_o. The transmission is higher for the smaller effective mass since a given number of holes (created by direct absorption

of the optical pulse) will more nearly fill the states available in the valence band up to the optically coupled states (see Fig. 1). Figure 3, which shows the transmission of a weak probe pulse as a function of time delay after the excitation pulse, illustrates the effect of m_o on the sample transmission as the carrier distribution cools.

The "best" choice for m_o probably lies between the value at the band edge and that calculated to give the correct experimental linear absorption coefficient at ω_o, since the former results in underestimating the available valence band states needed for saturation and the latter overestimates this number. Notice that, if a parabolic approximation is made to the energy band structure, a single effective mass cannot be chosen that accurately predicts both the linear absorption coefficient and the number of available states up to, and including the optically coupled states. Also note that the "best" choice for m_o at 300 K will be different from the "best" choice at 77 K since the direct band gap is narrower at 300 K, and the optically coupled states are located higher (lower) in the conduction (valence) band.

It is clear then that assuming a parabolic central conduction band valley leads to uncertainty in the value of the absolute transmission. Similar uncertainties are introduced by attempts to obtain a simplified model for the valence band and for the conduction band side valleys.

IV. PLASMONS, PHONONS AND PULSEWIDTH DEPENDENCE

In this section, we use the model of Elci, et al.[1] to investigate numerically the roles of the intravalley electron-phonon relaxation and plasmon-assisted electron-hole recombination in determining the dependence of the optical transmission on pulsewidth. As we pointed out in the introduction, the experimental measurements[5] are in substantial disagreement with the original calculations,[1] as shown in Fig. 4. Note that increasing the pulsewidth lowers the theoretical transmission for larger pulse intensities. However, there is little if any pulsewidth dependence in the data.

In the model, intravalley phonon-assisted relaxation and plasmon-assisted recombination will cool and heat, respectively, the carrier distribution. If the phonon-assisted cooling dominates, more of the electrons (holes) created by direct absorption of the optical pulse will be located lower (higher) in the conduction (valence) band. Thus the states needed for absorption will be more completely filled, and the transmission will be enhanced. However, if heating due to plasmon-assisted recombination is the stronger, the electrons (holes) will be located higher (lower) in the conduction (valence) band. The states needed for optical absorption will thus be freed, and this process will tend to decrease the transmission as a function of optical pulse energy.

To understand the pulsewidth dependence of these two processes, consider two pulses of equal energy but different width incident on a germanium sample. While the longer pulse interacts with the sample there is more time for both phonon-assisted relaxation and plasmon-assisted recombination. If heating due to the plasmon emission is dominant, the longer pulse will have a higher distribution temperature and, consequently, a lower transmission. This trend is illustrated in Fig. 5. This figure is a plot of the single pulse transmission of germanium as a function of incident pulse energy for the three pulse-widths: 2.5, 5, and 10 psec. The phonon-assisted relaxation terms are neglected in this figure. If, however, the phonon-assisted relaxation is stronger, similar arguments show that the transmission will have the opposite pulsewidth dependence, as presented in Fig. 6. Figure 6 is identical to Fig. 5 except that the plasmon-assisted recombination terms have been neglected while retaining the phonon relaxation terms.

The actual pulsewidth dependence is determined by which of these two processes is stronger. The coupling constants in the paper by Elci, et al.[1] were such that phonon-assisted cooling of the carrier distribution was small during the time the pulse was incident on the sample, thus plasmon heating effects determined the pulsewidth dependence. The experimental data obtained by Bessey, et al.[5] indicate that either the plasmon effects were overestimated in the original calculations

or the phonon-assisted relaxation was underestimated. Indeed, an uncertainty does exist in both processes.

Consider for example, the accuracy with which the electron-optical phonon coupling constant Q_o is known. Table 2 contains a list of several experimentally and theoretically determined values for Q_o. Evidently, there is a substantial uncertainty in Q_o. This uncertainty has a dramatic effect on the evolution of the optical transmission both while the excitation pulse is incident on the sample and after its passage. Figure 7 illustrates the change that occurs in the transmission as a function of incident pulse energy when Q_o is varied. The plasmon terms have the same strength as in Ref. 1. For the lower values of Q_o, there is negligible cooling during the period the pulse interacts with the sample; the final values of transmission are determined primarily by direct absorption followed by plasmon heating of the carrier distribution. For this range of values, the dependence on Q_o is small. However, for the larger values of Q_o (for example, $Q_o = 2\times10^{-3}\ \frac{erg}{cm}$), the phonon-assisted cooling competes with the plasmon-assisted heating causing the electrons (holes) to be located lower (higher) in the conduction (valence) band, which results in a higher transmission.

The uncertainty in Q_o also affects the temporal evolution of the germanium transmission after the passage of the excitation pulse as shown in Fig. 8. The rise in these

transmission curves is directly determined by the time it takes the electrons (holes) to relax to the bottom (top) of the conduction (valence) band and more completely clog the states needed for absorption. In fact, interrogation of the temporal evolution of the transmission of the sample by a weak probe pulse would represent a sensitive method for determining Q_o, if all other parameters associated with the model were precisely known.

There is also an uncertainty in the plasmon-assisted recombination that is a result of the approximations made in the calculation. In Ref. 1, the plasmon recombination rate was calculated using the Drude form of the dielectric function. Thus, the calculation of the plasmon rate must be regarded as phenomenological with the plasmon broadening being an estimated constant in this model. The plasmon broadening was estimated to be proportional to ten times the rate for scattering of an electron from the Γ-valley to the (111)-valley accompanied by the emission of a phonon.

One convenient way of demonstrating the effect of the uncertainty in the plasmon recombination rate is to vary the plasmon broadening by varying the phonon-assisted intervalley scattering time, τ_o, which is inversely proportional to the broadening. Figure 9 illustrates the change in the transmission as a function of incident pulse energy when τ_o is varied. Increasing τ_o markedly increases the transmission for

larger incident energies. For larger τ_o, the plasmon broadening is smaller and plasmon recombinations are less effective in heating the carrier distribution. Thus, the electrons (holes) are located lower (higher) in the conduction (valence) band; the transmission is higher.

If either the electron-optical phonon coupling constant or the plasmon broadening were known with certainty, the other constant could be estimated within this model by varying the unknown constant to obtain the proper transmission and pulsewidth dependence. However, the uncertainties in the phonon relaxation and plasmon recombination rates make an accurate evaluation of either constant impossible within the present model.

CONCLUSION

In the previous sections, we have illustrated the effect of the simplified band structure, the phonon-assisted relaxation, and the plasmon-assisted recombination on the transmission of picosecond pulses through germanium. In particular, we have shown that the pulsewidth dependence of the plasmon effects is opposite to that of the phonon-assisted relaxation. Because of the uncertainties in the electron-optical phonon coupling constant and the plasmon broadening, the deviation of the experimentally-measured pulsewidth dependence of the germanium transmission from the initially predicted pulsewidth dependence can be accounted for within the framework of the model, if the proper balance is maintained between plasmon recombination and phonon relaxation during the passage of the pulse through the sample. Since this pulsewidth dependence can be adjusted almost at will because of these uncertainties, the pulsewidth experiments do not provide a definitive test of the model. We believe the model to be essentially correct since it agrees with a substantial body of experimental data; however, further experimental and theoretical study is required to completely determine the model and its physical constants.

REFERENCES

1. A. Elci, M. O. Scully, A. L. Smirl and J. C. Matter, (to be published in July 1, 1977 issue of Phys. Rev. B.)
2. C. J. Kennedy, J. C. Matter, A. L. Smirl, H. Weichel, F. A. Hopf, and S. V. Pappu, Phys. Rev. Lett 32, 419 (1974).
3. A. L. Smirl, J. C. Matter, A. Elci, and M. O. Scully, Opt. Commun. 16, 118 (1976).
4. H. M. van Driel, J. S. Bessey, and R. C. Hanson, (to be published).
5. J. S. Bessey, B. Bosacchi, H. M. van Driel, and M. O. Scully, Bull. Am. Phys. Soc. 22, 379 (1977); J. S. Bessey, B. Bosacchi, H. M. van Driel, and A. L. Smirl, (submitted for publication).
6. See Section III of Reference 1.
7. See footnote 36 of Reference 1.
8. See Eq. 30 of Reference 1.

LIST OF TABLES

TABLE I

$\frac{m_o}{m_e}$	Optically Coupled Energy in Γ-valley $E_\Gamma(\hbar\omega_o)$ in (ev)	Optically Coupled Energy in Valence Band $E_h(\hbar\omega_o)$ in (ev)	Number of Electron States Available up to $E_\Gamma(\hbar\omega_o)$ in the Γ-valley n_o in (cm^{-3})	Number of Hole States Available up to $E_h(\hbar\omega_o)$ in the two Valence Bands n_h in (cm^{-3})	Number of Electron States Available up to $E_\Gamma(\hbar\omega_o)$ in the ten side valleys n_G in (cm^{-3})	Theoretical Beer's Law Absorption Coefficient from Eq. 30 in Ref. 1 α_o in (μm^{-1})
0.04	1.1316	0.0384	6.772×10^{18}	1.354×10^{19}	1.067×10^{21}	0.978
0.06	1.1153	0.0548	1.152×10^{19}	2.304×10^{19}	9.982×10^{20}	1.168
0.1	1.0870	0.0830	2.148×10^{19}	4.296×10^{19}	8.822×10^{20}	1.438
0.12	1.0748	0.0952	2.642×10^{19}	5.284×10^{19}	8.333×10^{20}	1.540

TABLE II

Electron-Optical Phonon Coupling Constant, Q_o (X 10^{-4} $\frac{erg}{cm}$)	Reference
7.68	K. Seeger, Semiconductor Physics (Springer-Verlag, Wien-New York, 1973), p. 213.
14.4	E. M. Conwell, Solid State Physics, edited by F. Seitz, D. Turnbull, and H. Ehrenreich, (Academic Press, New York, 1967), Suppl. 9, p. 171.
8.0	H. S. Reik and H. Risken, Phys. Rev. 126, 1737 (1962).
6.4	S. M. deVeer and H. J. G. Meyer, in Proc. 6th Intern. Conf. Phys. Semicond., Exeter, 358 (1962).
11.2	M. H. Jorgensen, N. I. Meyer, and K. J. Schmidt-Tiedemann, in Proc. 7th Intern. Conf. Phys. Semicond., Paris, 457 (1964).
10.72	W. Fawcett and E.G.S. Paige, J. Phys. C. 4, 1801 (1971).
18.5	H.J.G. Meyer, Phys. Rev. 112, 298 (1958).
8.16	D. C. Herbert, W. Fawcett, A. H. Lettington, and D. Jones in Proc. 11th Intern. Conf. Phys. Semicond., Warsaw, 1221 (1972).

LIST OF FIGURES

Fig. 7 The single pulse transmission of germanium as a function of incident pulse energy for several values of the electron-optical phonon coupling constant, Q_o.

Fig. 8 The instantaneous probe transmission as a function of time delay after an excitation pulse of photon density S_L (number of quanta/interaction volume) for several values of the electron-optical phonon coupling constant, Q_o.

Fig 9 The single pulse transmission of germanium as a function of incident pulse energy for several values of the plasmon broadening; τ_o ($\sim 10^{-14}$sec.) is the intervalley phonon-assisted scattering time used in the broadening in Ref. 1.

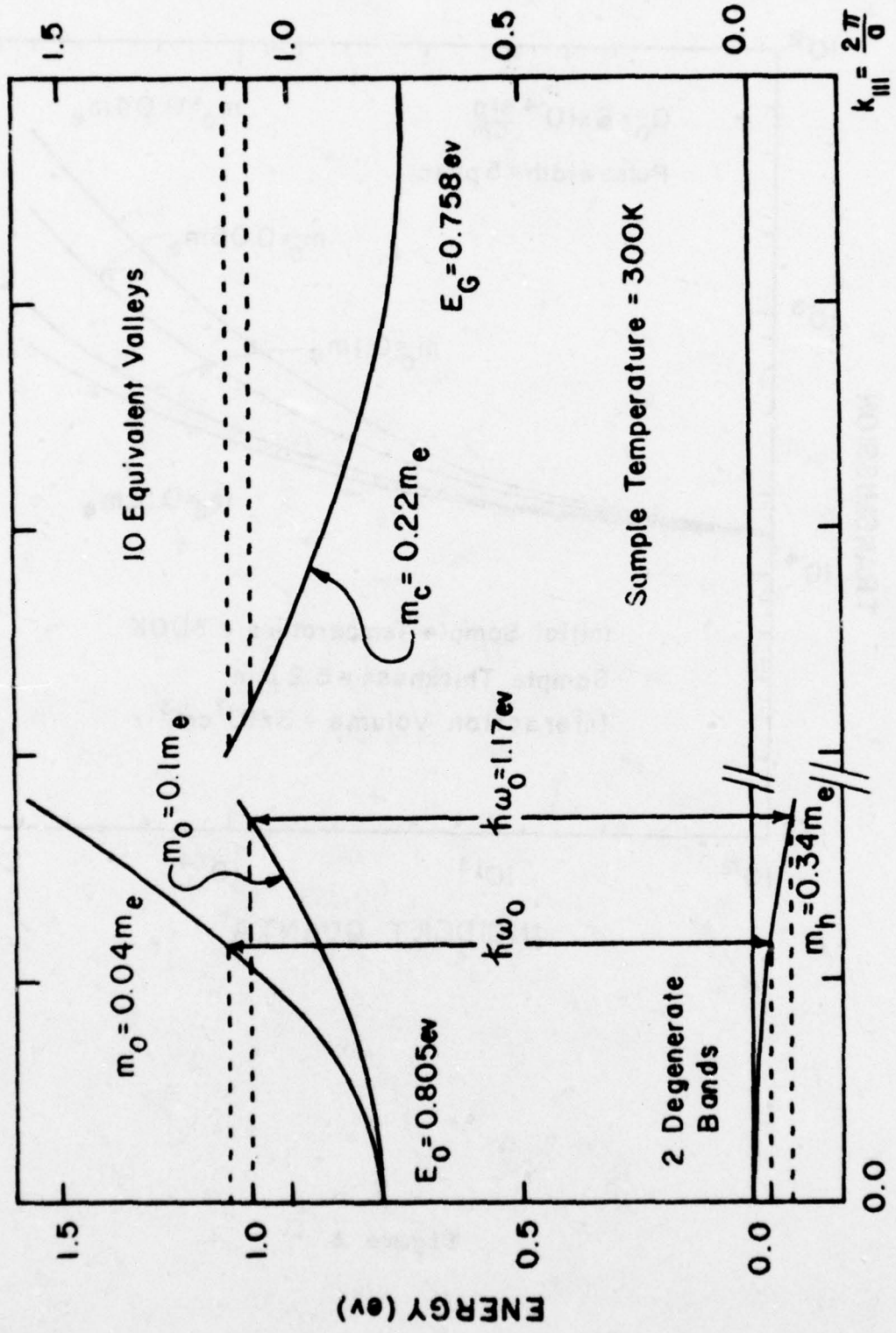

ENERGY (ev)
1.5
1.0
0.5
0.0
10 Equivalent Valleys
$m_o = 0.04 m_e$
$m_o = 0.1 m_e$
$m_c = 0.22 m_e$
$E_G = 0.758 ev$
$E_o = 0.805 ev$
$\hbar\omega_o$
$\hbar\omega_o = 1.17 ev$
2 Degenerate Bands
$m_h = 0.34 m_e$
Sample Temperature = 300K
WAVE NUMBER
$k_{111} = \frac{2\pi}{a}$

Figure 1

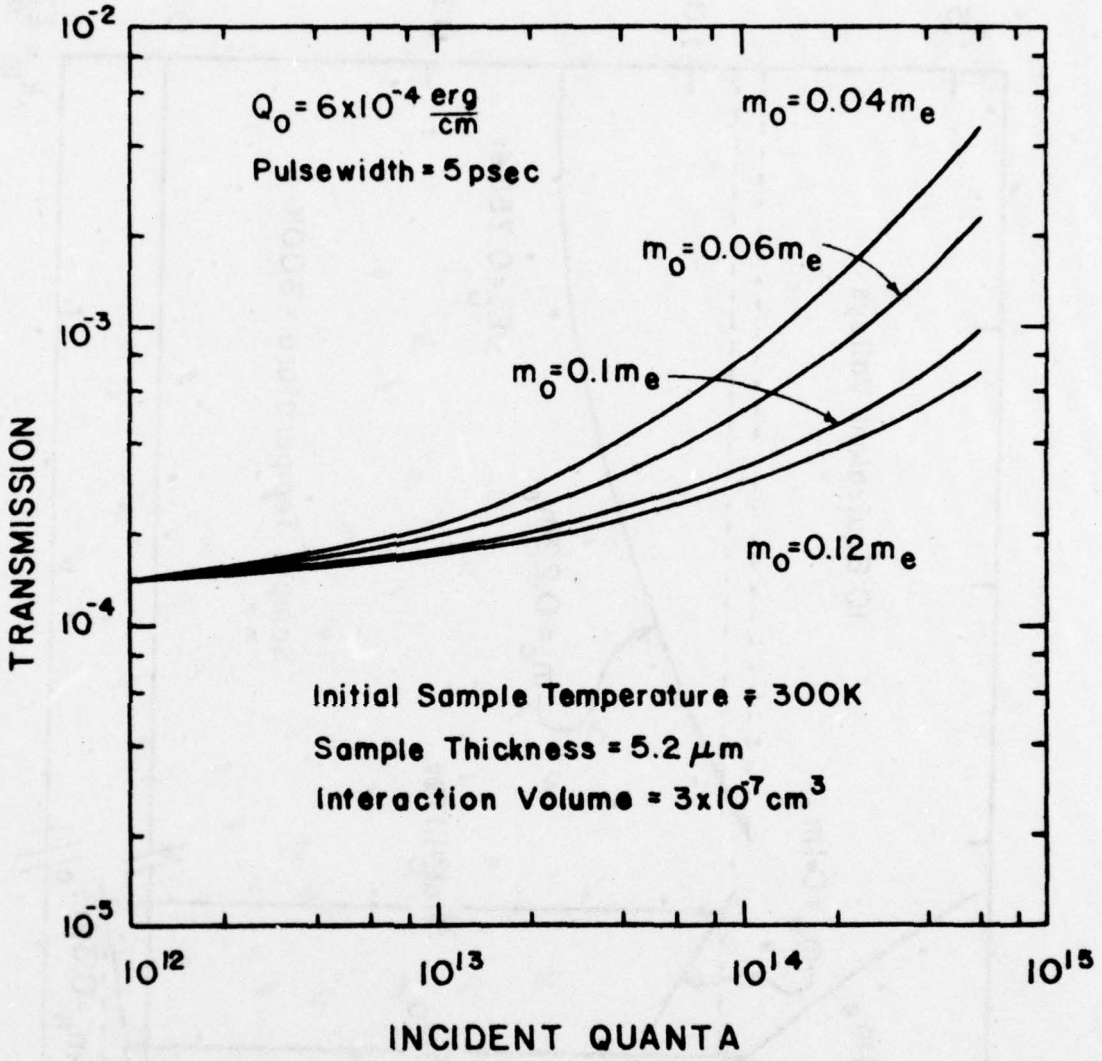

10^{-2}
10^{-3}
10^{-4}
10^{-5}
TRANSMISSION
$Q_0 = 6 \times 10^{-4} \frac{erg}{cm}$
Pulsewidth = 5 psec
$m_0 = 0.04 m_e$
$m_0 = 0.06 m_e$
$m_0 = 0.1 m_e$
$m_0 = 0.12 m_e$
Initial Sample Temperature = 300K
Sample Thickness = 5.2 μm
Interaction Volume = $3 \times 10^{-7} cm^3$
10^{12}
10^{13}
10^{14}
10^{15}
INCIDENT QUANTA

Figure 2

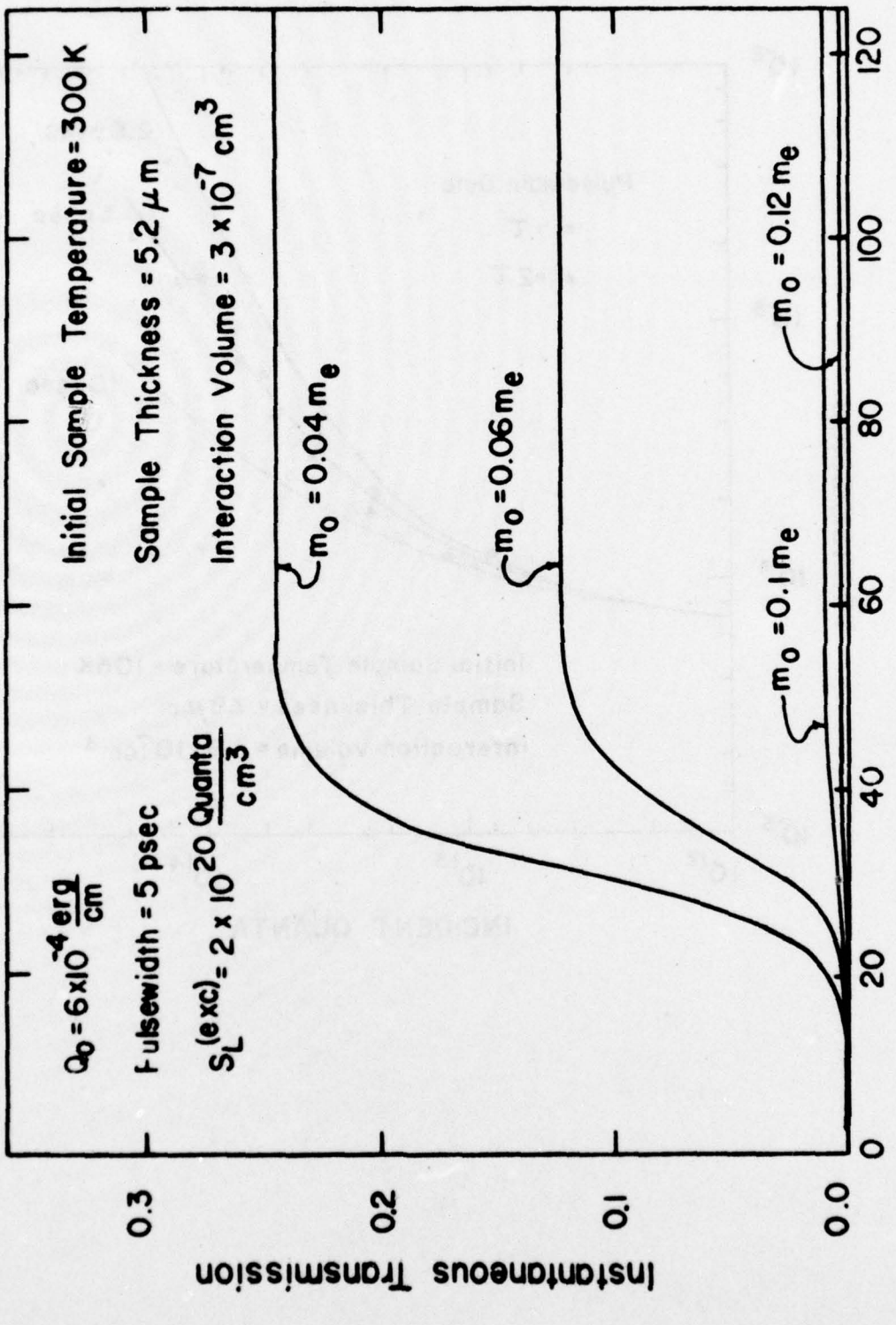

Initial Sample Temperature = 300 K
Sample Thickness = 5.2 μm
Interaction Volume = 3 x 10^-7 cm^3
Q0 = 6 x 10^-4 erg/cm
Pulsewidth = 5 psec
S_L^(exc) = 2 x 10^20 Quanta/cm^3
m0 = 0.04 me
m0 = 0.06 me
m0 = 0.1 me
m0 = 0.12 me
Instantaneous Transmission
Time (psec)
0.0
0.1
0.2
0.3
0
20
40
60
80
100
120

Figure 3

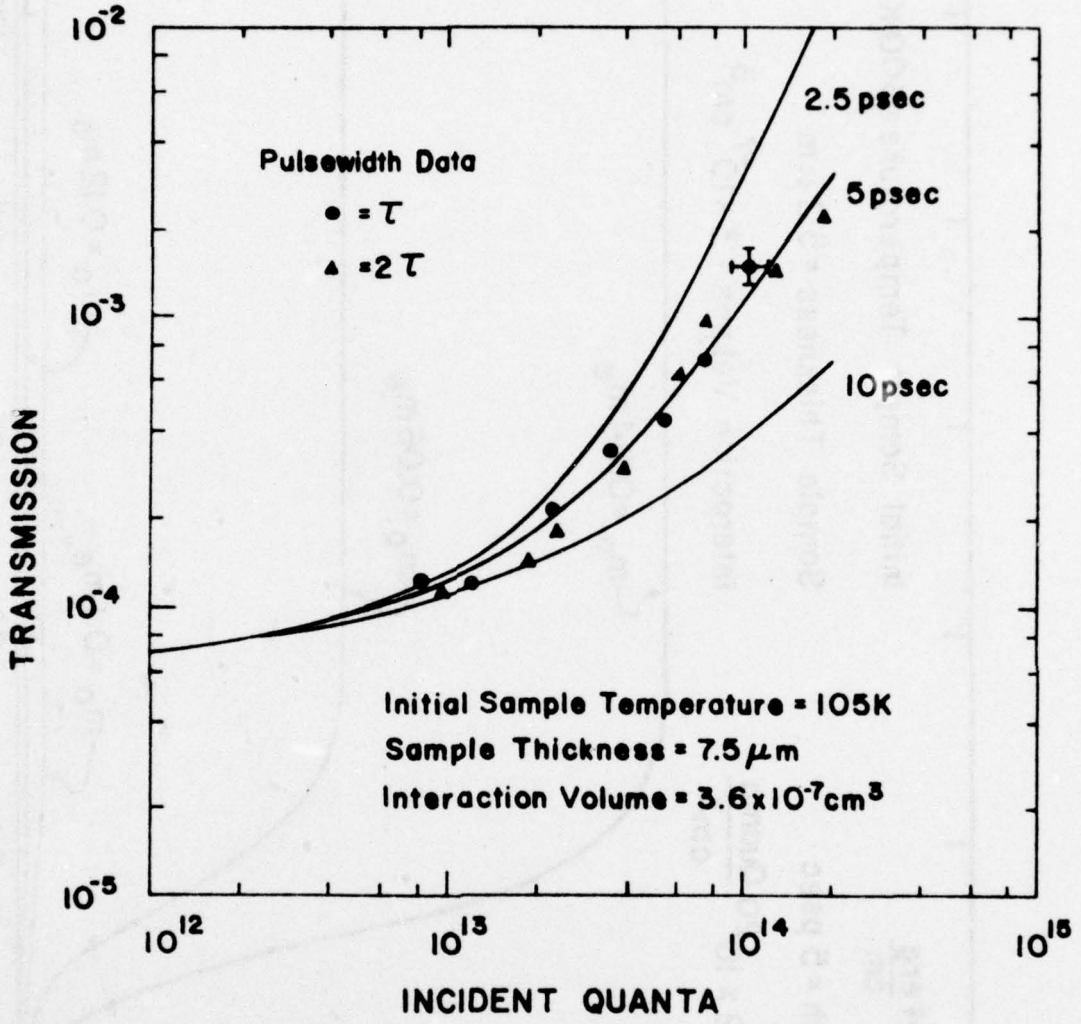

10^{-2}
10^{-3}
10^{-4}
10^{-5}
TRANSMISSION
2.5 psec
5 psec
10 psec
Pulsewidth Data
● = τ
▲ = 2τ
Initial Sample Temperature = 105K
Sample Thickness = 7.5 μm
Interaction Volume = $3.6 \times 10^{-7} cm^3$
10^{12}
10^{13}
10^{14}
10^{15}
INCIDENT QUANTA

Figure 4

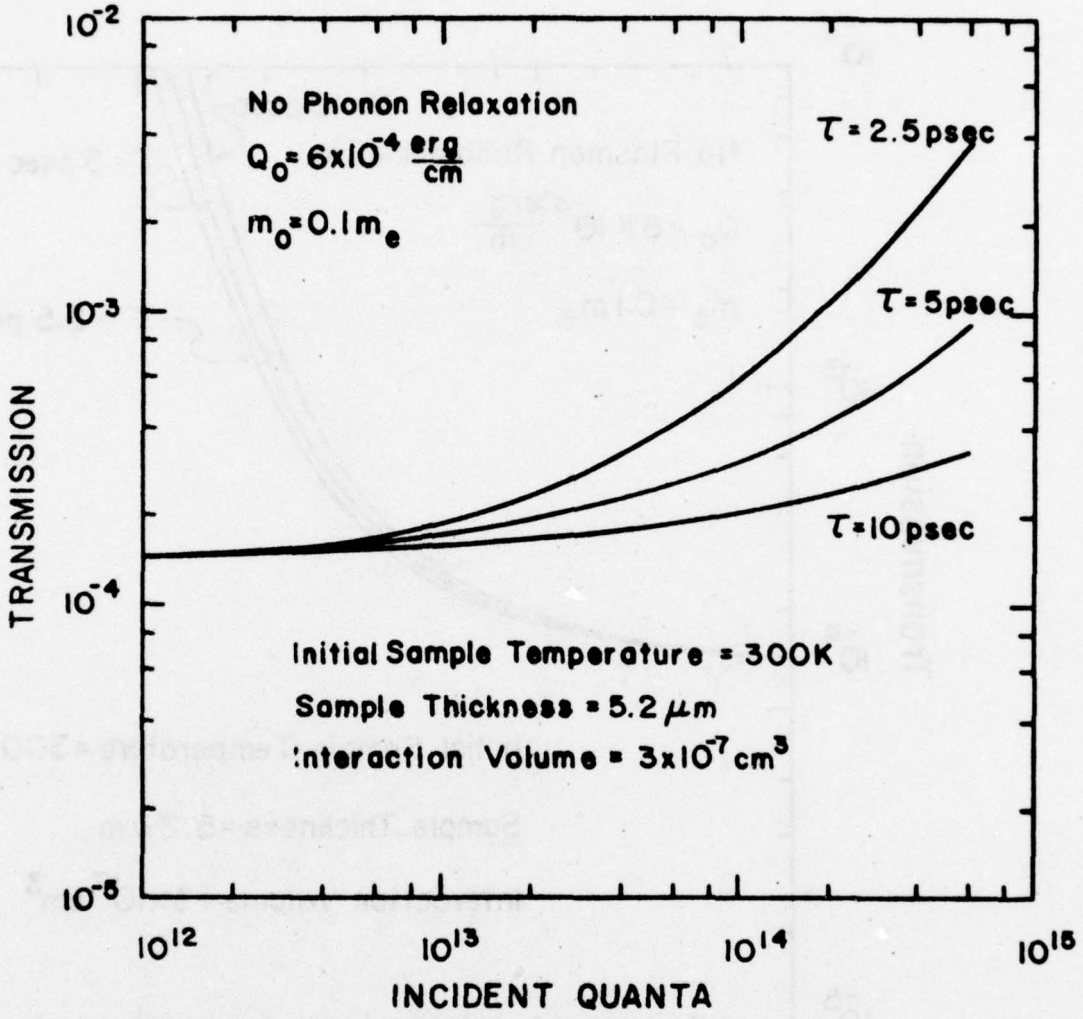

No Phonon Relaxation
$Q_0 = 6 \times 10^{-4} \frac{erg}{cm}$
$m_0 = 0.1 m_e$
$\tau = 2.5$ psec
$\tau = 5$ psec
$\tau = 10$ psec
Initial Sample Temperature = 300K
Sample Thickness = 5.2 μm
Interaction Volume = 3×10^{-7} cm^3
TRANSMISSION
INCIDENT QUANTA
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{12}
10^{13}
10^{14}
10^{15}

Figure 5

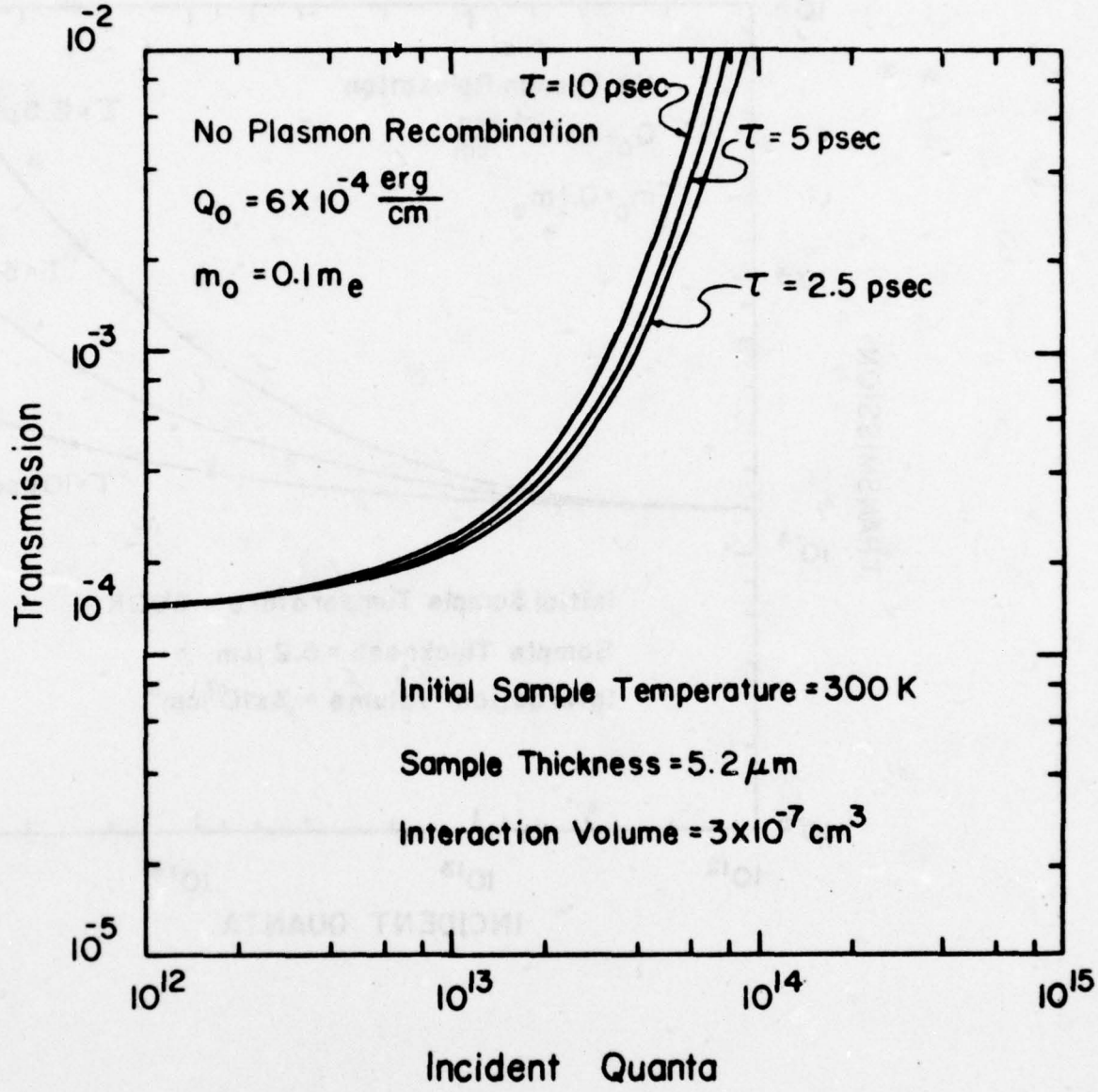

10^{-2}
10^{-3}
10^{-4}
10^{-5}
Transmission
τ = 10 psec
τ = 5 psec
τ = 2.5 psec
No Plasmon Recombination
$Q_0 = 6 \times 10^{-4} \frac{erg}{cm}$
$m_0 = 0.1 m_e$
Initial Sample Temperature = 300 K
Sample Thickness = 5.2 μm
Interaction Volume = 3×10^{-7} cm^3
10^{12}
10^{13}
10^{14}
10^{15}
Incident Quanta

Figure 6

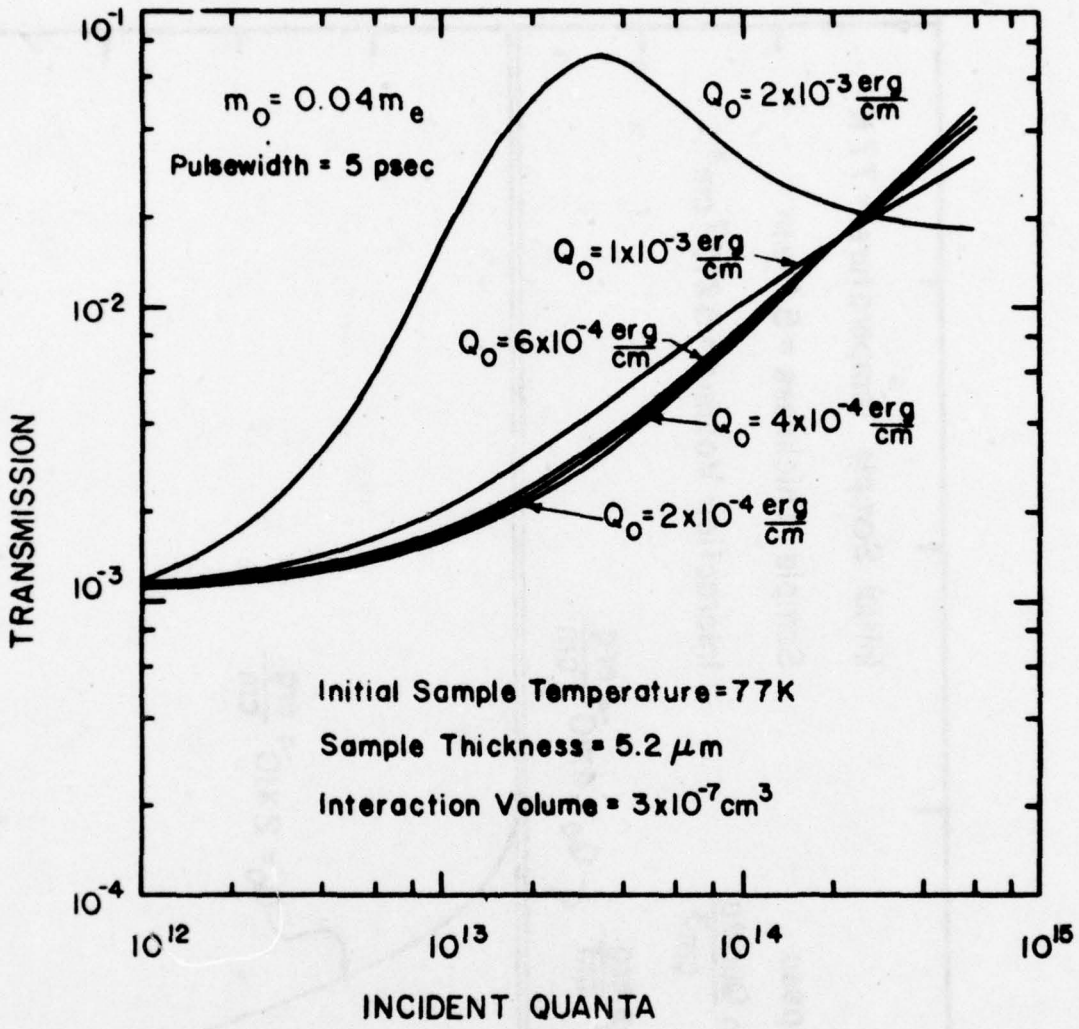

10^{-1}
10^{-2}
10^{-3}
10^{-4}
TRANSMISSION
$m_o = 0.04 m_e$
Pulsewidth = 5 psec
$Q_o = 2 \times 10^{-3} \frac{erg}{cm}$
$Q_o = 1 \times 10^{-3} \frac{erg}{cm}$
$Q_o = 6 \times 10^{-4} \frac{erg}{cm}$
$Q_o = 4 \times 10^{-4} \frac{erg}{cm}$
$Q_o = 2 \times 10^{-4} \frac{erg}{cm}$
Initial Sample Temperature = 77K
Sample Thickness = 5.2 μm
Interaction Volume = $3 \times 10^{-7} cm^3$
10^{12}
10^{13}
10^{14}
10^{15}
INCIDENT QUANTA

Figure 7

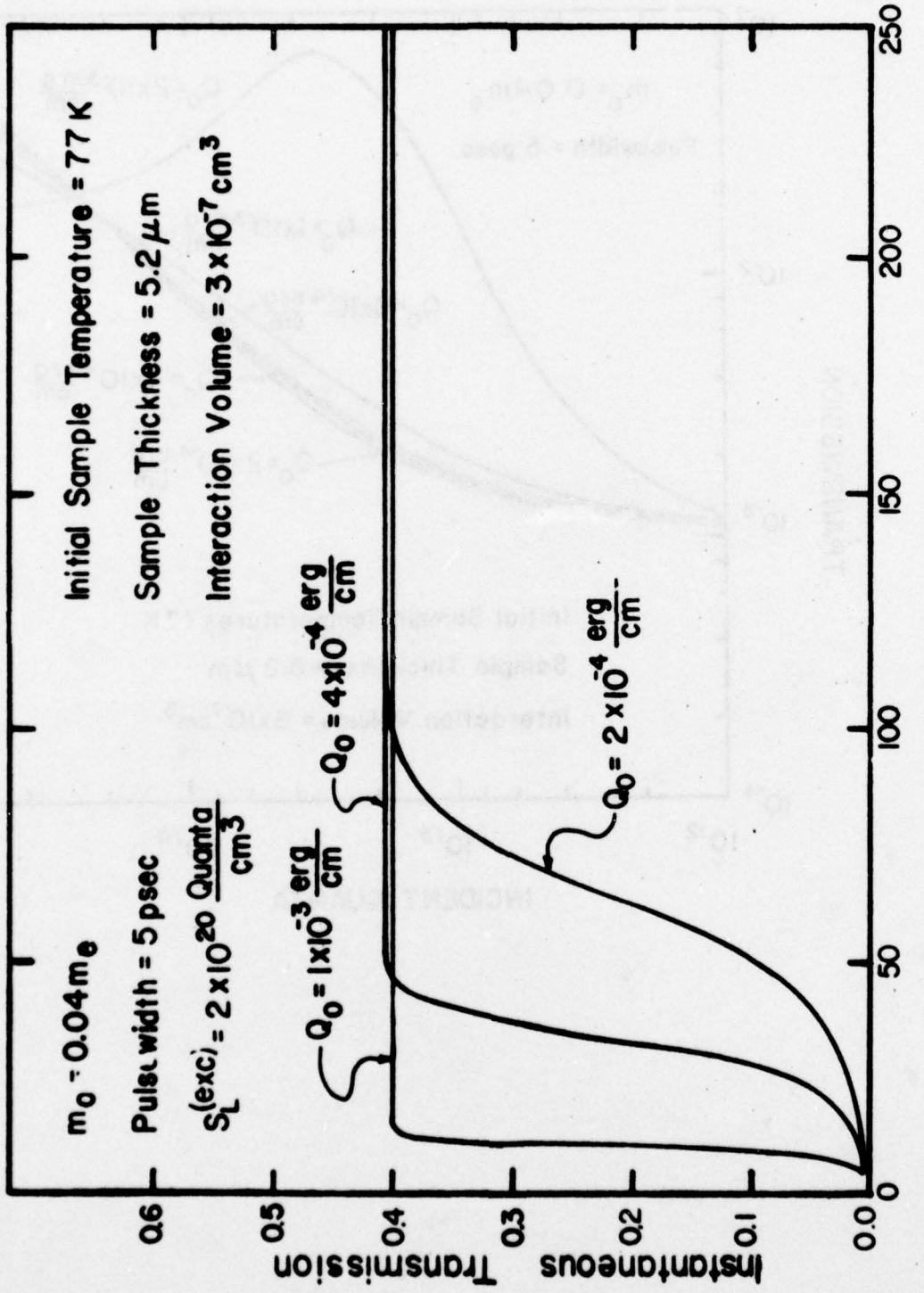

$m_o = 0.04 m_e$
Pulse width = 5 psec
$S_L^{(exc)} = 2 \times 10^{20} \frac{Quanta}{cm^3}$
Initial Sample Temperature = 77 K
Sample Thickness = 5.2 μm
Interaction Volume = $3 \times 10^{-7} cm^3$
$Q_o = 1 \times 10^{-3} \frac{erg}{cm}$
$Q_o = 4 \times 10^{-4} \frac{erg}{cm}$
$Q_o = 2 \times 10^{-4} \frac{erg}{cm}$
Instantaneous Transmission
Time (psec)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
50
100
150
200
250

Figure 8

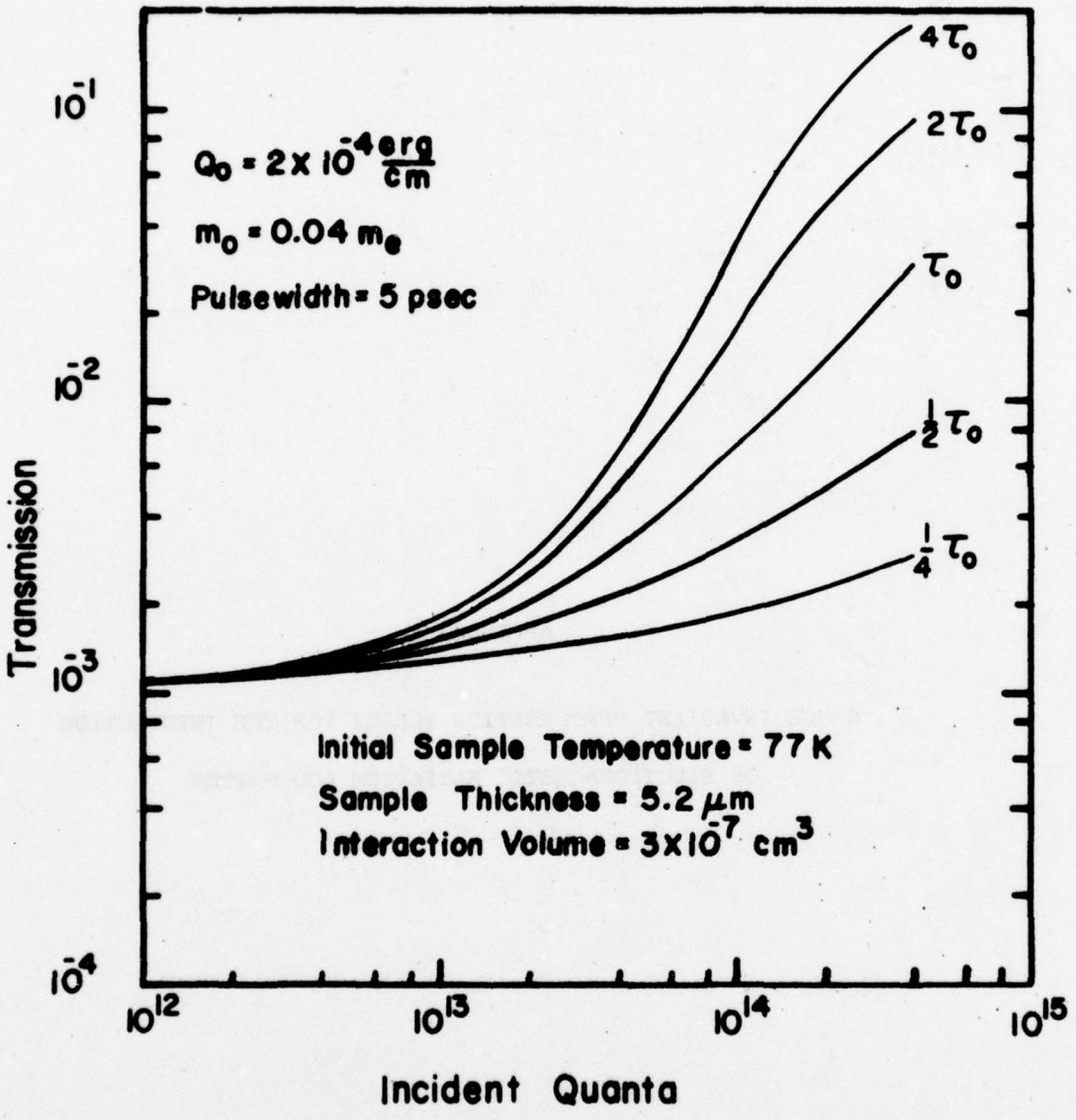

$4\tau_0$
$2\tau_0$
τ_0
$\frac{1}{2}\tau_0$
$\frac{1}{4}\tau_0$
$Q_0 = 2 \times 10^{-4} \frac{erg}{cm}$
$m_0 = 0.04\ m_e$
Pulsewidth = 5 psec
Initial Sample Temperature = 77 K
Sample Thickness = 5.2 μm
Interaction Volume = 3×10^{-7} cm^3
Transmission
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{12}
10^{13}
10^{14}
10^{15}
Incident Quanta

Figure 9

APPENDIX E

GAUGE INVARIANT PERTURBATION THEORY FOR THE INTERACTION OF ELECTROMAGNETIC RADIATION AND MATTER

Submitted to American Journal of Physics

Gauge Invariant Perturbation Theory for the Interaction of Electromagnetic Radiation and Matter*

Donald H. Kobe and Arthur L. Smirl
Department of Physics
North Texas State University
Denton, Texas 76203

*Supported in part by the Office of Naval Research and the North Texas State University Faculty Research Fund.

ABSTRACT

The conventional perturbation theory in quantum mechanics for the interaction of electromagnetic radiation with matter, which is based on the time-dependent vector and scalar potentials, is shown not to be gauge invariant. The perturbation problem is reformulated, using the approach of K.-H. Yang, in a gauge-invariant way, in which the transitions between states are induced by the quantum mechanical power operator. In order to use an unperturbed Hamiltonian which does not involve the electromagnetic field, the gauge is chosen to give the electric dipole interaction. This choice can be made if the magnetic field is negligible and the electric field is slowly varying over atomic dimensions. As an example, the Kramers-Heisenberg formula is derived from the gauge invariant electric dipole interaction.

I. INTRODUCTION

Over half a century has elapsed since the discovery of quantum mechanics and the first treatment of radiation problems in the new quantum theory.[1] Although most quantum mechanics texts treat the interaction of radiation and matter, few books do more than mention gauge transformations and relegate them to the problems.[2-5] The necessity of doing calculations in a gauge invariant way is often not emphasized. In fact, many books proceed to expand out the terms involving the vector potential, and treat them as perturbations.[6-9] Other books use the electric dipole interaction from the start, without necessarily relating it to the vector potential.[10,11] A few books do point out that the electric dipole interaction is related to the vector potential interaction by a gauge transformation.[12-16] The perturbing question arises, however, as to which interaction to use - the one involving the vector potential or the electric dipole interaction - or is it merely a matter of taste? It is the purpose of this paper to show that it does make a difference and the electric dipole interaction should be used when it can be. It may be necessary to use higher order electric multipoles and magnetic multipoles as well.

In order to substantiate this claim, a gauge invariant time-dependent perturbation theory is developed.[17] Since the Schrödinger equation can not usually be solved exactly, the interaction between the radiation and the atom[18] is treated by perturbation theory. The most convenient and natural choice for the unperturbed Hamiltonian is the Hamiltonian for the atom in the absence of the radiation field. As we will show, this choice of the unperturbed Hamiltonian implies that the electric dipole interaction is the correct gauge-invariant interaction to use in these problems, when the magnetic multipole and higher-order electric multipole interactions are negligible.

The reason the electric dipole interaction for these problems plays such a fundamental role is that it is the only interaction for which the atomic Hamiltonian and the electromagnetic field are decoupled in a gauge-invariant way. Thus the expansion coefficients of the wave function in terms of the unperturbed atomic states can be interpreted as true, gauge-independent probability amplitudes for finding the atomic system in an unperturbed state. For another choice of gauge, the gauge-invariant unperturbed Hamiltonian is shown to couple the atom and the electromagnetic field. If the Hamiltonian is broken up in a way that does not

preserve gauge invariance, the expansion coefficients for the wave function are gauge dependent and cannot be interpreted as probability amplitudes. The fundamental significance of the electric dipole interaction was noted by Lamb[19] in 1952 when he remarked that another choice of gauge gave results that were not in agreement with experiment.

The subject of gauge transformations in quantum mechanics has been recently discussed in an excellent article by Yang,[17] who emphasizes the importance of doing calculations in a gauge invariant way. Some of his points are reviewed here for completeness, and to make the fundamental ideas available to a wider audience. In particular, Yang uses the correspondence principle to show that the gauge-invariant energy operator for the system is not the Hamiltonian, which is gauge dependent, but the Hamiltonian without the time-dependent scalar potential. From this fundamental point, a gauge-invariant time-dependent perturbation theory is developed, in which the transitions between states is induced by the gauge-invariant quantum mechanical power operator. We have extended his treatment by writing the power operator in terms of the electric dipole energy and derivatives of the electric field, which shows the naturalness and convenience of the electric dipole approximation. The gauge invariant perturbation theory is then applied to the scattering of light by an atom to derive the Kramers-Heisenberg formula.[20] These ideas can be used to

supplement the usual textbook treatment of radiation theory in quantum mechanics and quantum optics.

The role of gauge transformations in quantum theory was apparently first pointed out by Fock.[21] Weyl[22] developed a unified theory of electromagnetism and matter based on a principle of gauge invariance. His theory was the harbinger of the modern gauge theories of particle physics.[23] The principle of gauge invariance states that the Schrödinger equation is form invariant under simultaneous gauge transformation of the wave function and the vector and scalar potentials. Subsequently, Pauli[24] called the gauge transformation of the wave function a "gauge transformation of the first kind," while the usual gauge transformation of the electromagnetic potentials he called a "gauge transformation of the second kind."

The goal of developing a completely gauge invariant formulation of quantum electrodynamics has been pursued by many workers,[25] and has been beset with many technical difficulties.[26] A resurgence of interest in the role of the electromagnetic potentials in quantum theory occurred when Aharonov and Bohm[27] suggested that they had some significance apart from the electromagnetic fields they described. The development of the laser also stimulated interest in the interaction of the electromagnetic field with atoms, molecules, and solids. The gauge invariant Hamiltonian for a system of nonrelativistic particles interacting with the radiation field was derived by Power and

Zienau[28] using a canonical transformation. This method has been used and extended by many others.[29] A good review of this work has recently been given by Woolley.[30]

In Sec.II the conventional perturbation treatment of radiation is reviewed to point out some common misconceptions. In Sec. III the form invariance of the Schrödinger equation under gauge transformations is shown. The gauge invariance of expectation values for observables and their equations of motion is discussed in Sec. IV, which leads to the gauge invariant energy operator. The formulation of a gauge-invariant time-dependent perturbation theory is given in Sec. V . Then in Sec. VI the electric dipole interaction is derived by making a particular choice of gauge and neglecting magnetic and higher-order electric multipoles. In Sec. VII these ideas are applied to derive the Kramers-Heisenberg formula for the scattering of light by an atom. Finally the paper is summarized and conclusions given in Sec VIII.

II. CONVENTIONAL APPROACH

The conventional treatment of radiation problems in quantum mechanics[6-8] is reviewed here briefly to point out its inadeqacies and some common misconceptions. The main objection is that the conventional perturbation theory is not gauge invariant. This fact leads to possible spurious interactions and expansion coefficients that are not probability amplitudes. For an unperturbed Hamiltonian which does not depend on the electromagnetic

field, the gauge should be chosen such that the electric dipole interaction is the perturbation, which is discussed in Sec. VI. This interaction is not a weak field approximation to the conventional treatment, but is the correct gauge invariant interaction when the magnetic field is negligible and the electric field is slowly varying over the atom.

The Hamiltonian for a particle of mass m, charge q, and momentum $\underset{\sim}{p} = -i\hbar\nabla$ in an electromagnetic field described by the time varying vector potential $\underset{\sim}{A}$ and scalar potential A_o is

$$H = H(\underset{\sim}{A},A_o) = (\underset{\sim}{p} - q\underset{\sim}{A}/c)^2/2m + V + qA_o \ , \qquad (2.1)$$

where V is an external static potential energy. The quadratic term in Eq. (2.1) can be expanded to give

$$H(\underset{\sim}{A},A_o) = H_o - (q/2mc)(\underset{\sim}{A}\cdot\underset{\sim}{p} + \underset{\sim}{p}\cdot\underset{\sim}{A}) + (q^2/2mc^2)A^2 + qA_o, \qquad (2.2)$$

where H_o is the Hamiltonian H(0,0) in the absence of an electromagnetic field

$$H_o = p^2/2m + V \ . \qquad (2.3)$$

The Coulomb gauge

$$\nabla\cdot\underset{\sim}{A} = 0 \qquad (2.4)$$

is often used in Eq. (2.2) so that $\underset{\sim}{p}\cdot\underset{\sim}{A} = \underset{\sim}{A}\cdot\underset{\sim}{p}$. The conventional procedure is to use H_o as the unperturbed Hamiltonian, and the remainder in Eq. (2.2) as the perturbation. However, this approach is not gauge invariant. The coefficients in the expansion of the wave function ψ in terms of the eigenstates of H_o are gauge dependent, and can therefore not be interpreted as probability amplitudes. Appendix A gives a discussion of the lack of gauge invariance of the expansion coefficients in the conventional perturbation theory. Even the Hamiltonian in Eq. (2.1) or (2.2) is not gauge invariant, which is discussed in Sec. IV.

That there are difficulties with the conventional perturbation theory can be seen by considering the case in which there is no electromagnetic field. The situation in which the electric field $\underset{\sim}{E} = 0$ and the magnetic induction $\underset{\sim}{B} = 0$ can be described by the potentials $\underset{\sim}{A} = 0$ and $A_o = 0$. Then the perturbation in Eq. (2.2) is zero, as it should be. However, as we show in Sec. III, the same situation of no electromagnetic field can also be described by the vector potential $\underset{\sim}{A} = \nabla\Lambda$ and the scalar potential $A_o = -\partial\Lambda/\partial(ct)$, where Λ is an arbitrary function of space and time. Then the perturbation in Eq. (2.2) is not zero, but is a spurious interaction. When treated by perturbation theory, this interaction would give incorrect results. It is horrifying to think of the time that could be spent in calculating the

effects of this spurious interaction to nth order. This example should emphasize the importance of doing perturbation theory in a gauge invariant way.

As we will see in Sec. VI, the gauge can be chosen, in problems where the magnetic field is negligible and the wavelength of the radiation is long compared to atomic dimensions, such that Eq. (2.1) becomes

$$H = H_o - q\underset{\sim}{E}(0,t)\cdot\underset{\sim}{r}, \tag{2.5}$$

where $\underset{\sim}{E}(0,t)$ is the electric field at the origin (chosen to be the center of the atom). The electric dipole interaction in Eq. (2.5) is obviously gauge invariant, since it does not depend on the potentials. It is also the correct perturbation to use in conjunction with H_o.

It has sometimes been thought that the Hamiltonian in Eq. (2.5) is only an approximation to Eq. (2.2) that is applicable in the weak field case when the A^2 term is negligible,[31] because the electric dipole interaction is linear in the electric field. This erroneous conclusion stems from the similarity in form between the $-q\underset{\sim}{E}\cdot\underset{\sim}{r}$ term and the $-(q/mc)\underset{\sim}{A}\cdot\underset{\sim}{p}$ term in Eq. (2.2) when the Coulomb gauge in Eq. (2.4) is used.

The matrix elements of the two interaction terms are not equal, but they are, however, related. If, for example, A_o

is chosen to be zero (as it can be in the absence of charge density), then the electric field is

$$\underset{\sim}{E} = -\partial\underset{\sim}{A}/\partial(ct) \quad \text{for} \quad A_o = 0 \ . \tag{2.6}$$

For a harmonically varying field of frequency ω, Eq. (2.6) implies that

$$\underset{\sim}{E}(0,t) = i\omega\underset{\sim}{A}(0,t)/c \ . \tag{2.7}$$

The commutation relations between the components x_i of $\underset{\sim}{r}$ and p,

$$[x_i, p_j] = i\hbar\, \delta_{ij}, \tag{2.8}$$

where $i = 1,2,3$, implies that the commutator between H_o in Eq. (2.3) and $\underset{\sim}{r}$ is

$$[H_o, \underset{\sim}{r}] = -i\hbar\underset{\sim}{p}/m \tag{2.9}$$

The eigenstates ϕ_n of H_o satisfy the eigenvalue problem

$$H_o\phi_n = e_n\, \phi_n \ , \tag{2.10}$$

where e_n is the eigenenergy. When matrix elements of Eq. (2.9) are taken between the eigenstates ϕ_n, the result is

$$\omega_{nm}\ (\underset{\sim}{r})_{nm} = -i(\underset{\sim}{p})_{nm}/m \ , \tag{2.11}$$

where ω_{nm} is defined as

$$\hbar\,\omega_{nm} = e_n - e_m \quad , \tag{2.12}$$

and the matrix elements are defined as

$$(\underset{\sim}{r})_{nm} = \langle \phi_n | \underset{\sim}{r} \phi_m \rangle \quad . \tag{2.13}$$

The matrix elements of the electric dipole interaction are thus

$$-q\,\underset{\sim}{E}(0,t)\cdot(\underset{\sim}{r})_{nm} = (\omega/\omega_{nm})\;(-q/mc)\underset{\sim}{A}(0,t)\cdot(\underset{\sim}{p})_{nm} \tag{2.14}$$

when Eqs. (2.7) and (2.11) are used.

If the resonance condition,

$$\hbar\omega = e_n - e_m \tag{2.15}$$

is satisfied, it appears that Eq. (2.5) is equivalent to Eq. (2.2) without the A^2 and A_o terms.[32] However, for virtual processes in perturbation theory, the resonance condition in Eq. (2.15) is not satisfied. This point will be discussed in more detail in Sec. VII.

Since the electric dipole interaction in Eq. (2.5) and the $-(q/mc)\,A\cdot p$ interaction in Eq. (2.2) give matrix elements differing by a factor of ω/ω_{nm} in Eq. (2.14), it can be decided experimentally which interaction to use. Lamb[33] has

shown experimentally that the electric dipole interaction of Eq. (2.5) is the proper interaction.

We shall show in Sec. VI on theoretical grounds that the electric dipole interaction is not a weak field approximation to the full Hamiltonian in Eq. (2.2), but is fully equivalent to Eq. (2.1) under the approximations that (a) magnetic effects are negligible (i.e., the magnetic induction $\underset{\sim}{B}$ does not appear) and (b) the field $\underset{\sim}{E}(\underset{\sim}{r},t)$ is slowly varying over the atom (i.e., the dipole approximation is valid). Furthermore, Eq. (2.5) is the correct Hamiltonian to use when H_o is taken to be the unperturbed Hamiltonian. It is interesting to note that many early workers in quantum theory did use the electric dipole approximation.[20,34] In the next section the form invariance of the Schrödinger equation under gauge transformations is discussed.

III. GAUGE TRANSFORMATIONS

The Schrödinger equation for a particle in an electromagnetic field described by the vector potential $\underset{\sim}{A}$ and the scalar potential A_o is shown in this section to be form invariant if the wave function ψ and the potentials are transformed in an appropriate way. The electromagnetic field is unchanged if the potentials undergo a gauge transformation, while the probability density $\psi^*\psi$ is unchanged if ψ is multiplied by a phase factor. These transformations were shown

by Fock[21] as early as 1927 to be intimately related. Discussions of the form invariance of the Schrödinger equation under gauge transformations are available in several books,[35,37] but the subject is reviewed here because of its importance for subsequent sections.

The Schrödinger equation for a particle of mass m in an electromagnetic field is

$$(1/2m)\ (\underset{\sim}{p}-q\underset{\sim}{A}/c)^2\ \psi + V\psi + qA_o\psi = i\hbar\ \partial\psi/\partial t \qquad (3.1)$$

when the Hamiltonian of Eq. (2.1) is used. A new wavefunction ψ' can be obtained by multiplying ψ by a phase factor,

$$\psi' = \exp(i\ q\Lambda/c\hbar)\psi\ , \qquad (3.2)$$

where $\Lambda = \Lambda(\underset{\sim}{r},t)$ is a real function. Equation (3.2) is called a local gauge transformation of the first kind after Pauli.[24]

If Eq. (3.2) is substituted into Eq. (3.1) the Schrödinger equation becomes

$$(1/2m)\ (\underset{\sim}{p}-q\underset{\sim}{A}'/c)^2\ \psi' + V\psi' + qA'_o\psi' = i\hbar\partial\psi'/\partial t\ , \qquad (3.3)$$

which has the same mathematical form as Eq. (3.1). In order to preserve the form, the new vector potential $\underset{\sim}{A}'$ is defined as

$$\underset{\sim}{A}' = \underset{\sim}{A} + \nabla\Lambda\ , \qquad (3.4)$$

and the new scalar potential A_o' as

$$A_o' = A_o - \partial\Lambda/\partial(ct) \quad . \tag{3.5}$$

Equations (3.4) and (3.5) are the customary gauge transformations of electromagnetic theory,[38] which are called gauge transformations of the second kind by Pauli.[24]

The nonuniqueness of the vector and scalar potentials is due to the way in which the observed electric and magnetic fields are expressed in terms of the potentials.[38] The magnetic induction

$$\underset{\sim}{B} = \nabla \times \underset{\sim}{A} \quad , \tag{3.6}$$

is invariant under Eq. (3.4). Likewise, the electric field

$$\underset{\sim}{E} = -\nabla A_o - \partial \underset{\sim}{A}/\partial(ct) \quad , \tag{3.7}$$

is also invariant under Eqs. (3.4) and (3.5). Therefore any choice of gauge function Λ can be used and the same electric and magnetic fields will be obtained. However, the wave function is changed according to Eq. (3.2).

Thus the Schrödinger equation is form invariant under the local gauge transformation of the first kind in Eq. (3.2) if the scalar and vector potentials transform according to the customary gauge transformations of the second kind in Eqs. (3.4) and (3.5). In the next section the gauge invariance of expectation values for observables is discussed.

IV. GAUGE INVARIANCE OF EXPECTATION VALUES

The expectation value of a physical observable should not depend on the choice of gauge. If it did, then its value would be arbitrary, since the gauge function Λ in Sec. III is arbitrary. In this section, the consequences of this reasonable requirement are investigated. Following Yang,[17] the correspondence principle is used to determine the proper operators for observables.

If θ is a Hermitian operator which describes an observable, then its expectation value in the state ψ must be equal to its expectation value in the state ψ' in Eq. (3.2),

$$\langle\psi|\theta\psi\rangle = \langle\psi'|\theta'\psi'\rangle \quad , \tag{4.1}$$

in order to ensure gauge invariance. The gauge transformed operator θ' is defined as

$$\theta' = \exp(iq\Lambda/\hbar c)\ \theta\ \exp(-iq\Lambda/\hbar c), \tag{4.2}$$

which will be called a gauge transformation of the first kind on the operator. A prime denotes a gauge transformation of the first kind in Eq. (4.2) on all operators <u>except</u> the vector potential $\underline{A}$ and the scalar potential A_o, for which a prime denotes a gauge transformation of the second kind defined in Eqs. (3.4) and (3.5), respectively. Under a gauge transformation

of the first kind in Eq. (4.2) $\underset{\sim}{A}$ and A_o are unchanged, so the prime on them is reserved for the gauge transformation of the second kind.

The Schrödinger equation in Eq. (3.1) is form invariant under the gauge transformation on the wave function in Eq. (3.2), when Eqs. (3.4) and (3.5) are used for the new potentials. The expectation values of all operators are independent of gauge transformations of the first kind when Eqs. (4.1) and (4.2) are used. However, not all operators are form invariant under the gauge transformation of the first kind in Eq. (4.2). For example, the momentum operator $\underset{\sim}{p} = -i\hbar\nabla$ is not form invariant under Eq. (4.2), but the combination $(\underset{\sim}{p} - q\underset{\sim}{A}/c)$ is.[39] A gauge transformation of the first kind in Eq. (4.2) on $(\underset{\sim}{p} - q\underset{\sim}{A}/c)$ gives

$$\begin{aligned}(\underset{\sim}{p} - q\underset{\sim}{A}/c)' &= \exp(iq\Lambda/\hbar c)(\underset{\sim}{p} - q\underset{\sim}{A}/c)\exp(-iq\Lambda/\hbar c)\\ &= (\underset{\sim}{p} - q\nabla\Lambda/c) - q\underset{\sim}{A}/c\\ &= (\underset{\sim}{p} - q\underset{\sim}{A}'/c),\end{aligned} \tag{4.3}$$

where Eq. (3.4) has been used for $\underset{\sim}{A}'$. The calculation here has been given explicitly to show how the vector potential becomes transformed by a gauge transformation of the second kind as a consequence of the gauge transformation of the first kind on $\underset{\sim}{p}$. Likewise, the operator $i\hbar\,\partial/\partial t$ is not form invariant under the gauge transformation in Eq. (4.2), but the combination $(i\hbar\,\partial/\partial t - qA_o)$ is. A gauge transformation of the first kind on

$(i\hbar\partial/\partial t - qA_o)$ gives

$$(i\hbar\partial/\partial t - qA_o)' = \exp(iq\Lambda/\hbar c)(i\hbar\partial/\partial t - qA_o)\exp(-iq\Lambda/\hbar c)$$
$$= (i\hbar\partial/\partial t + q\partial\Lambda/\partial(ct))-qA_o$$
$$= i\hbar\partial/\partial t - qA_o' \quad , \tag{4.4}$$

where Eq. (3.5) has been used for A_o'. The scalar potential becomes transformed by a gauge transformation of the second kind as a consequence of the gauge transformation of the first kind on $i\hbar\partial/\partial t$. We emphasize again that the prime on the vector potential $\underset{\sim}{A}'$ and the scalar potential A_o' denotes the gauge transformation of the second kind given in Eqs. (3.4) and (3.5), respectively. The prime on all other operators θ' denotes the gauge transformation of the first kind in Eq. (4.2). The operator θ may involve $\underset{\sim}{A}$ or A_o as in Eqs. (4.3) or (4.4), but the prime on the operator θ still denotes the operation shown in Eq. (4.2).

Another operator of particular interest that is <u>not</u> form invariant under Eq. (4.2) is the Hamiltonian $H(\underset{\sim}{A},A_o)$ in Eq. (2.1). The Hamiltonian transforms under Eq. (4.2) as

$$H'(\underset{\sim}{A},A_o) = \exp(iq\Lambda/\hbar c)\, H(\underset{\sim}{A},A_o)\exp(-iq\Lambda/\hbar c)$$
$$= H(\underset{\sim}{A}',A_o)$$
$$= H(\underset{\sim}{A}',A_o') + q\partial\Lambda/\partial(ct) \quad , \tag{4.5}$$

where Eqs. (4.3), (3.4), and (3.5) have been used. Although

the time evolution of the system with wave function ψ is described by the Hamiltonian $H(\underset{\sim}{A},A_o)$ through the Schrödinger equation, it need not represent the instantaneous energy operator. In fact Eq. (4.5) shows that the total Hamiltonian $H(\underset{\sim}{A},A_o)$ can not represent the energy operator, since its expectation value is not gauge invariant, i.e.,

$$\begin{aligned} \langle\psi|H(\underset{\sim}{A},A_o)\psi\rangle &= \langle\psi'|H'(\underset{\sim}{A},A_o)\psi'\rangle \\ &= \langle\psi'|H(\underset{\sim}{A}',A_o)\psi'\rangle \\ &\neq \langle\psi'|H(\underset{\sim}{A}',A_o')\psi'\rangle \quad . \end{aligned} \tag{4.6}$$

The question thus arises as to the proper choice for the energy operator and other operators representing physical observables. This question is answered in the remainder of this section.

An operator representing a physical observable must, of course, have an expectation value which is independent of the choice of gauge. Yang[17] has used the correspondence principle to select the appropriate operators to represent physical observables. According to him, an operator represents a physical quantity with a classical analogue only if the equation of motion for the expectation value of the operator is of the same form as the equation of motion for the corresponding classical quantity.

The equation of motion for the expectation value of an operator θ can be obtained by differentiating Eq. (4.1) with respect to the time t, and using the Schrödinger equation. The Schrödinger equation in Eq. (3.3) can be written as

$$H(\underset{\sim}{A}',A_o')\psi' = i\hbar\partial\psi'/\partial t \quad , \tag{4.7}$$

where $H(\underset{\sim}{A}',A_o')$ is the Hamiltonian in Eq. (2.1) with the vector and scalar potentials $\underset{\sim}{A}'$ and A_o', respectively. The time derivative of Eq. (4.1) can be written as

$$\begin{aligned} i\hbar d\langle\psi'|\theta'\psi'\rangle/dt = &-\langle\psi'|H(\underset{\sim}{A}',0)\theta'\psi'\rangle \\ &+\langle\psi'|(i\hbar\partial/\partial t - qA_o')\theta'\psi'\rangle \quad , \end{aligned} \tag{4.8}$$

where the scalar potential has been taken out of the Hamiltonian. From Eqs. (2.1) and (4.5) it can be seen that

$$H(\underset{\sim}{A}',0) = (\underset{\sim}{p} - q\underset{\sim}{A}'/c)^2/2m + V = H'(\underset{\sim}{A},0) \tag{4.9}$$

is form invariant under gauge transformation. Thus when Eqs. (4.9), (4.4), (4.2) and (4.1) are used, it can be seen that Eq. (4.8) is manifestly gauge invariant, i.e., it is valid if the primes are removed.

Even though Eq. (4.8) is manifestly gauge invariant, it is not the most convenient form in which to write the time development of the expectation value. The time derivative in the last term of Eq. (4.8) can be carried out to give

$$\begin{aligned} i\hbar d\langle\psi'|\theta'\psi'\rangle/dt = &\langle\psi'|[\theta',H(\underset{\sim}{A}',A_o')]\psi'\rangle \\ &+ \langle\psi'|(i\hbar\partial\theta'/\partial t)\psi'\rangle \quad . \end{aligned} \tag{4.10}$$

The manifest gauge invariance has been lost in this traditional form,[40] since the right hand side of Eq. (4.10) is the sum of two terms, neither of which is gauge invariant. However, since the left hand side of Eq. (4.10) is gauge invariant, the sum of the two terms on the right hand side gives a gauge invariant result.

Equation (4.10) can now be applied to give the time dependence of some observables, for which the explicit gauge invariance can be seen by inspection. Since Eq. (4.10) is gauge invariant, the primes can be removed. When it is applied to the position operator $\underset{\sim}{r}$, the time rate of change of the average particle position is[41]

$$m d\langle\psi|\underset{\sim}{r}\psi\rangle/dt = \langle\psi|(\underset{\sim}{p} - q\underset{\sim}{A}/c)\psi\rangle \quad . \tag{4.11}$$

By using Eqs. (4.1) and (4.3) it can be seen that Eq. (4.11) is manifestly gauge invariant, i.e., it is valid with primes on ψ, $\underset{\sim}{r}$, and $\underset{\sim}{A}$. From the classical analogy the operator on the right hand side of Eq. (4.11) is used to define the velocity operator

$$\underset{\sim}{v} = m^{-1}(\underset{\sim}{p} - q\underset{\sim}{A}/c) , \tag{4.12}$$

which is form invariant under gauge transformation of the first kind.

The time rate of change of the average velocity operator

$\underset{\sim}{v}$ in Eq. (4.12) can also be obtained by applying Eq. (4.10) (without the primes). The equation of motion is[41]

$$\langle\psi|\underset{\sim}{v}\psi\rangle/dt = -\langle\psi|(\nabla V)\psi\rangle + \langle\psi|q\underset{\sim}{E}\psi\rangle + (1/2c)\langle\psi|(\underset{\sim}{v}\times\underset{\sim}{B} - \underset{\sim}{B}\times\underset{\sim}{v})\psi\rangle \quad , \tag{4.13}$$

where the electric field $\underset{\sim}{E}$ and the magnetic induction $\underset{\sim}{B}$ are defined in Eqs. (3.7) and (3.6), respectively. Equation (4.13) is the quantum mechanical analogue of Newton's second law, since the right hand side is the average of the forces. The last term in Eq. (4.13) is the average value of the Hermitian Lorentz force operator.

The energy of the system should also be a gauge invariant quantity.[17] The question was previously raised in Eq.(4.6) as to what operator to use for the energy. The Hamiltonian $H(\underset{\sim}{A},A_o)$ is not form invariant under gauge transformation as Eq. (4.5) shows. However, Eq. (4.9) shows that $H(\underset{\sim}{A},0)$ is form invariant under a gauge transformation of the first kind. If Eq. (4.10) is applied to $H(\underset{\sim}{A},0)$, the result is

$$d\langle\psi|H(\underset{\sim}{A},0)\psi\rangle/dt = \langle\psi|P\psi\rangle \quad . \tag{4.14}$$

The right hand side of Eq. (4.14) is the average of the quantum mechanical power operator P defined as

$$P = (q/2)(\underset{\sim}{E}\cdot\underset{\sim}{v} + \underset{\sim}{v}\cdot\underset{\sim}{E}) \quad , \tag{4.15}$$

where the velocity operator $\underset{\sim}{v}$ is defined in Eq. (4.12). Equation (4.14) is also manifestly gauge invariant, i.e., it is valid with primes on ψ, $\underset{\sim}{A}$, and P. From Eq. (4.13) it can be seen that the electric force is $q\underset{\sim}{E}$. The electric force $q\underset{\sim}{E}$ does work, but not the magnetic force. Equation (4.15) is the quantum mechanical version of the classical statement that the power put into the system is the time rate of change of the energy of the system. We can therefore identify $H(\underset{\sim}{A},0)$ with the quantum mechanical energy operator.

V. GAUGE INVARIANCE OF PROBABILITY AMPLITUDES

In this section a gauge invariant form of time-dependent perturbation theory is developed, following the method of Yang.[17] The customary perturbation theory, discussed in Sec. II, is based on using H_o in Eq. (2.3) as the unperturbed Hamiltonian, and the remainder of Eq. (2.2) as the perturbation. As shown in Appendix A, the customary approach is not gauge invariant. For an arbitrary gauge, the interpretation of the expansion coefficients in the wave function as probability amplitudes for finding the system in an energy eigenstate of the unperturbed Hamiltonian H_o is not correct.

As Lamb[19] has noted, the correct gauge to choose to interpret the expansion coefficients as probability amplitudes of finding the system in eignstates of the unperturbed Hamiltonian H_o is the one for which the Hamiltonian has the electric dipole interaction

given in Eq. (2.5). This choice of gauge, however, is not arbitrary, but is seen to be a consequence of the gauge invariant formulation of time-dependent perturbation theory.

The energy operator $H(\underset{\sim}{A},0)$, shown to be form invariant under gauge transformations in the last section, is used as the unperturbed Hamiltonian. It satisfies the eigenvalue problem

$$H(\underset{\sim}{A},0)\psi_n(t) = \varepsilon_n(t)\psi_n(t) \quad , \tag{5.1}$$

where ψ_n and ε_n are in general functions of the time t because $\underset{\sim}{A}$ is time dependent. The wave function ψ for the Schrödinger equation in Eq. (3.1) can be expanded in terms of the complete set of states $\{\psi_n\}$ as

$$\psi = \sum_n c_n\psi_n \quad . \tag{5.2}$$

When Eq. (5.2) is substituted into Eq. (3.1) and Eq. (5.1) is used, the equation of motion for c_n is

$$i\hbar\dot{c}_n = \varepsilon_n c_n + \sum_m \langle\psi_n|(qA_o - i\hbar\partial/\partial t)\psi_m\rangle c_m \quad . \tag{5.3}$$

In contrast to Eq. (A2), Eq. (5.3) is gauge invariant. It should be emphasized that ψ_m in Eq. (5.3) is in general a function of time because of Eq. (5.1).

It is now shown that the coefficients c_n in Eq. (5.2) are unchanged under a gauge transformation. If a gauge transformation of the first kind in Eq. (3.2) is made on the wave function ψ in the Schrödinger equation in Eq. (3.1), and the gauge transformations of Eqs. (3.4) and (3.5) are used, the transformed Schrödinger equation in Eq. (3.3) is obtained. The same procedure can be used on the eigenvalue problem in Eq. (5.1) to obtain

$$H(\underset{\sim}{A}',0)\psi_n' = \varepsilon_n\psi_n' \quad , \tag{5.4}$$

where ψ_n' is related to ψ_n by Eq. (3.2) and where Eq. (4.9) has been used. The unperturbed energy ε_n is unchanged under the gauge transformation on the wavefunction. Since both ψ' and ψ_n' are related to ψ and ψ_n by Eq. (3.2), the expansion of ψ' in terms of ψ_n' has the same expansion coefficients as Eq. (5.2),

$$\psi' = \sum_n c_n\psi_n' \quad . \tag{5.5}$$

Thus, when Eq. (5.5) is substituted into the Schrödinger equation in Eq. (3.3), the equation for the time-dependent coefficients c_n is

$$i\hbar\dot{c}_n = \varepsilon_n c_n + \sum_m \langle\psi_n' \mid (qA_o' - i\hbar\partial/\partial t)\psi_m'\rangle c_m \quad . \tag{5.6}$$

From Eqs. (4.4), (4.2), and (4.1) it can be seen that the matrix

elements in Eq. (5.6) are equal to the matrix elements in Eq. (5.3). Since $\{c_n\}$ are independent of the choice of gauge they can be interpreted as the probability amplitude of finding the system in energy eigenstates of $H(\underset{\sim}{A},0)$, the physically meaningful unperturbed Hamiltonian.

Although Eq. (5.3) has been shown to be gauge invariant, its physical significance can be clarified by rewriting the interaction term. If Eq. (5.1) is differentiated with respect to time it can be shown (Appendix B) that for $n \neq m$

$$\langle\psi_n|(qA_o - i\hbar\partial/\partial t)\psi_m\rangle = i\hbar(\varepsilon_n - \varepsilon_m)^{-1}\langle\psi_n|P\psi_m\rangle \qquad (5.7)$$

where P is the power operator in Eq. (4.15). Therefore Eq. (5.3) can be written as

$$(i\hbar\partial/\partial t - \tilde{\varepsilon}_n)\, c_n = \sum_{m\neq n} i\hbar(\varepsilon_n - \varepsilon_m)^{-1}\langle\psi_n|P\psi_m\rangle\, c_m \quad , \qquad (5.8)$$

where the "dressed" eigenenergy $\tilde{\varepsilon}_n$ is defined as

$$\tilde{\varepsilon}_n = \varepsilon_n + \langle\psi_n|(qA_o - i\hbar\partial/\partial t)\psi_n\rangle \quad . \qquad (5.9)$$

Equation (5.8) remains gauge invariant because the matrix elements of the power operator P in Eq. (4.15) are gauge invariant. The dressed eigenenergy $\tilde{\varepsilon}_n$ in Eq. (5.9) is also gauge invariant because

both terms on the right hand side are gauge invariant by Eqs. (5.1), (4.4), (4.2) and (4.1). Equation (5.8) explicitly shows that the transitions between states are induced by the power operator. In contrast, in the conventional time-dependent perturbation theory in Eq. (A2) the transitions between states are induced by gauge dependent potentials.

Although Eq. (5.8) shows clearly the physical significance of the interaction which induces transitions, it can be rewritten in a form in which the electric dipole approximation becomes more natural. The classical power can be written as

$$P_{cl} = q\underset{\sim}{E}\cdot\underset{\sim}{v} = d(q\underset{\sim}{E}\cdot\underset{\sim}{r})/dt - q(\partial\underset{\sim}{E}/\partial t)\cdot\underset{\sim}{r} - q(\partial E_i/\partial x_j)x_i v_j \qquad \text{(classical)}, \tag{5.10}$$

in terms of the components x_i of $\underset{\sim}{r}$, etc. The summation convention on repeated indices (i,j = 1,2,3) is used. The dipole energy $q\underset{\sim}{E}\cdot\underset{\sim}{r}$ explicitly appears in Eq. (5.10). An expression similar to Eq. (5.10) also holds quantum mechanically. In Appendix C it is shown that the matrix elements of the power operator are

$$\langle\psi_n|P\psi_m\rangle = (i\hbar)^{-1}\langle\psi_n|[q\underset{\sim}{E}\cdot\underset{\sim}{r},\ H(\underset{\sim}{A},0)]\psi_m\rangle - (q/4)\langle\psi_n|\{\{v_j,\partial E_i/\partial x_j\},x_i\}\psi_m\rangle \quad . \tag{5.11}$$

The curly brackets denote the anticommutator

$$\{a,b\} = ab + ba \quad , \tag{5.12}$$

which is used in Eq. (5.11) to give a Hermitian operator. Equation (5.11) is a quantum mechanical form of Eq. (5.10).

If Eq. (5.11) is substituted into Eq. (5.8), the result is

$$(i\hbar\partial/\partial t - \tilde{\varepsilon}_n)c_n = -\sum_{m\neq n}\Big|\langle\psi_n|q\underset{\sim}{E}\cdot\underset{\sim}{r}\psi_m\rangle + (i\hbar q/4)\,(\varepsilon_n-\varepsilon_m)^{-1}\langle\psi_n|\{\{v_j,\partial E_i/\partial x_j\},x_i\}\psi_m\rangle\Big|c_m \quad . \tag{5.13}$$

This equation is still exact and is also manifestly gauge invariant, since c_n, ε_n and the matrix elements are explicitly gauge invariant. Since the electric field $\underset{\sim}{E} = \underset{\sim}{E}(r,t)$ is a function of $\underset{\sim}{r}$, it cannot be taken outside of the inner product in Eq. (5.13). In the next section it will be shown how Eq. (5.13) can be simplified by making some approximations.

VI. ELECTRIC DIPOLE APPROXIMATION

If the magnetic field in the problem of interest is negligible, then we can simplify Eq. (5.13) by choosing the gauge such that the vector potential is zero. If we can further assume that the electric field is slowly varying over the dimensions of the atom, we obtain the electric dipole approximation.

Specifically, we choose the gauge function Λ in Eq. (3.2) such that the new vector potential $\underset{\sim}{A}'$ is,

$$\underset{\sim}{A}' = \underset{\sim}{A} + \nabla\Lambda = 0 \quad . \tag{6.1}$$

Thus the gauge function Λ is

$$\Lambda(\underset{\sim}{r},t) = - \int_0^{\underset{\sim}{r}} d\underset{\sim}{r}_1 \cdot \underset{\sim}{A}(\underset{\sim}{r}_1,t) + \Lambda(0,t), \tag{6.2}$$

which is path independent. The new scalar potential A_o' in Eq. (3.5) can be written as

$$\begin{aligned} A_o'(\underset{\sim}{r},t) = & \int_0^{\underset{\sim}{r}} d\underset{\sim}{r}_1 \cdot [\nabla_1 A_o(\underset{\sim}{r}_1,t) + \partial \underset{\sim}{A}(\underset{\sim}{r}_1,t)/\partial(ct)] \\ & + [A_o(0,t) - \partial\Lambda(0,t)/\partial(ct)] \end{aligned} \tag{6.3}$$

If Eq. (3.7) is substituted into Eq. (6.3) we have

$$A_o'(\underset{\sim}{r},t) = - \int_0^{\underset{\sim}{r}} d\underset{\sim}{r}_1 \cdot \underset{\sim}{E}(\underset{\sim}{r}_1,t) \tag{6.4}$$

if $\Lambda(0,t)$ is also chosen such that

$$\partial\Lambda(0,t)/\partial(ct) = A_o(0,t) \quad . \tag{6.5}$$

The Schrödinger equation in Eq. (3.3) thus becomes in the new gauge

$$H_o\psi' - q \int_0^{r'} d\underset{\sim}{r}_1 \cdot \underset{\sim}{E}(\underset{\sim}{r}_1,t)\psi' = i\hbar\partial\psi'/\partial t \quad , \tag{6.6}$$

where H_o in Eq. (2.3) is the Hamiltonian in the absence of the field. A multipole expansion of Eq. (6.6) can be made, but if the electric field is only slowly varying over atomic dimensions the main contribution is the electric dipole interaction,

$$[H_o - q\,\underset{\sim}{E}(0,t) \cdot \underset{\sim}{r}]\psi' = i\hbar\partial\psi'/\partial t \quad , \tag{6.7}$$

for which the Hamiltonian is the same as Eq. (2.5). Thus the contention that the two Hamiltonians in Eqs. (2.1) and (2.5) are equivalent if there are no magnetic effects and the dipole approximation is made is verified. The choice of gauge in Eqs. (6.1) and (6.4) makes Eq. (5.13) more tractable.

Equation (5.13) simplifies considerably if $\underset{\sim}{E}$ varies slowly over the extent of the wave functions ψ_n. In this electric dipole approximation $\underset{\sim}{E}$ can be replaced by its value at the origin $\underset{\sim}{E}(0,t)$, and the spatial derivatives of $\underset{\sim}{E}$ can be neglected. Then Eq. (5.13) can be written

$$(i\hbar\partial/\partial t - \tilde{\varepsilon}_n)c_n$$

$$= - \sum_{m\neq n} q\, \underset{\sim}{E}(0,t)\cdot \langle\psi_n|\underset{\sim}{r}\psi_m\rangle c_m, \qquad (6.8)$$

which is still valid for any choice of gauge.

This equation can be simplified still further if magnetic effects are negligible so that the gauge in which $\underset{\sim}{A} = 0$ and A_o is given by Eq. (6.4) can be used. Then the unperturbed Hamiltonian in Eq. (5.1) is $H(0,0) = H_o$ given in Eq. (2.3), for which the eigenvalue problem is given in Eq. (2.10) with $e_n = \varepsilon_n$. The scalar potential A_o in Eq. (6.4) becomes in the dipole approximation

$$A_o = -q\, \underset{\sim}{E}(0,t)\cdot \underset{\sim}{r} \quad . \qquad (6.9)$$

Thus, Eq. (6.8) can be rewritten as

$$(i\hbar\partial/\partial t - \varepsilon_n)c_n = \sum_{m\neq n} -q\, \underset{\sim}{E}(0,t)\cdot (\underset{\sim}{r})_{nm}\, c_m \ , \qquad (6.10)$$

where the matrix elements $(\underset{\sim}{r})_{nm}$ are defined in Eq. (2.13) and are now time independent. Equation (6.10) is completely equivalent to the Schrödinger equation in Eq. (6.7).

Thus the Hamiltonian of Eq. (2.5) is the physically meaningful Hamiltonian to use if H_o is chosen as the unperturbed Hamiltonian. The probability amplitudes and energies calculated will then be gauge invariant, in contrast

to the ones calculated from Eq. (2.2) using the Coulomb gauge in Eq. (2.4). In the next section these ideas will be applied to the scattering of a photon by an electron bound in an atom.

VII. KRAMERS-HEISENBERG FORMULA

In this section the equation for the time-dependent probability amplitudes derived in Eq. (6.10) is applied to calculate the scattering cross section for a photon from an atomic electron, which is called the Kramers-Heisenberg (KH) formula.[20] In the conventional perturbation theory[42] based on the Hamiltonian in Eq. (2.2), the three Feynman diagrams shown in Fig. 1 must be evaluated to obtain the KH formula to first order in the fine structure constant $e^2/\hbar c$. Even though the conventional perturbation theory is not gauge invariant, the KH formula is. A simplification can be made by using the gauge invariant formulation of the last section to calculate the cross section. Then only the two Feynman diagrams in Fig. 1(a) and (b) need to be evaluated if the $-\underset{\sim}{E}\cdot\underset{\sim}{r}$ interaction is used, where the charge $q = -e$.

In order to calculate the cross section for the scattering of photons, the quantized electromagnetic field is convenient.[43] The vector potential for the quantized electromagnetic field in the Coulomb gauge is[44]

$$\underset{\sim}{A}(\underset{\sim}{r},t) = \sum_{k,\alpha} c(\hbar/2\omega V)^{1/2} \{a_{\underset{\sim}{k}\alpha} \hat{\underset{\sim}{\varepsilon}}^{(\alpha)} \exp[i(\underset{\sim}{k}\cdot\underset{\sim}{r}-\omega t)] + \text{h.c.}\}, \quad (7.1)$$

where h.c. denotes the Hermitian conjugate of the preceding term, V is the volume, $\omega = kc$ is the angular frequency, and $A_o = 0$. The Coulomb gauge in Eq. (2.4) is used, and the polarization vectors $\hat{\underset{\sim}{\varepsilon}}^{(\alpha)}$ (α=1,2) are orthogonal to the wave vector $\underset{\sim}{k}$. The photon creation and annihilation operators, $a^{\dagger}_{\underset{\sim}{k}\alpha}$ and $a_{\underset{\sim}{k}\alpha}$, respectively, satisfy the boson commutation relations,

$$\begin{aligned} [a_{\underset{\sim}{k}\alpha},\ a^{\dagger}_{\underset{\sim}{k}'\alpha'}] &= \delta_{\underset{\sim}{k}\underset{\sim}{k}'}\,\delta_{\alpha\alpha'}\ , \\ [a_{\underset{\sim}{k}\alpha},\ a_{\underset{\sim}{k}'\alpha'}] &= 0\ , \\ [a^{\dagger}_{\underset{\sim}{k}\alpha},\ a^{\dagger}_{\underset{\sim}{k}'\alpha'}] &= 0\ . \end{aligned} \quad (7.2)$$

The quantized electric field $\underset{\sim}{E}$ is calculated from Eq. (2.6), which gives

$$\underset{\sim}{E}(\underset{\sim}{r},t) = \sum_{k,\alpha} (\hbar\omega/2V)^{1/2} \{ i a_{\underset{\sim}{k}\alpha} \hat{\underset{\sim}{\varepsilon}}^{(\alpha)} \exp[i(\underset{\sim}{k}\cdot\underset{\sim}{r}-\omega t)] + \text{h.c}\}. \quad (7.3)$$

The electric field is, of course, gauge invariant and can be used in Eq. (6.10).

In the initial state i the electron is in the atomic eigenstate ϕ_n of H_o in Eq. (2.10) with energy ε_n, while the photon has the wave vector $\underset{\sim}{k}$ and polarization α with energy $\hbar\omega$. In

the final state f the electron is in the state ϕ_m with energy ε_m, while the photon has the wave vector $\underset{\sim}{k}'$ and polarization α' with energy $\hbar\omega'$. The intermediate state in Fig. 1 (a) has one electron in the state ϕ_j with energy ε_j and no photons. In contrast, the intermediate state in Fig. 1 (b) has one electron in the state ϕ_j and two photons with wave vectors and polarization $(\underset{\sim}{k},\alpha)$ and $(\underset{\sim}{k}',\alpha')$. The total energy of the intermediate state in Fig. 1 (b) is $\varepsilon_j + \hbar\omega + \hbar\omega'$.

Equation (6.10) can be solved by iteration to second order to obtain the probability amplitude $c_m^{(2a)}(t)$ corresponding to Fig. 1 (a),

$$c_m^{(2a)}(t) = (ie^2/\hbar) \sum_j \frac{(\underset{\sim}{E}_o' \cdot \underset{\sim}{r})_{mj} (\underset{\sim}{E}_o \cdot \underset{\sim}{r})_{jn}}{\varepsilon_j - (\varepsilon_n + \hbar\omega + i0)} \times \int_{-T}^{'} dt_1 \exp(i\Delta E t_1/\hbar) \exp(i\varepsilon_n T/\hbar - i\varepsilon_m t/\hbar) \tag{7.4}$$

The electric field amplitude $\underset{\sim}{E}_o$ is defined as

$$\underset{\sim}{E}_o = (\hbar\omega/2V)^{1/2} \hat{\underset{\sim}{\varepsilon}} \quad , \tag{7.5}$$

and $\underset{\sim}{E}_o'$ is defined analogously with ω and $\hat{\underset{\sim}{\varepsilon}}$ replaced by ω' and $\hat{\underset{\sim}{\varepsilon}}'$ respectively. Equation (7.3) for $\underset{\sim}{E}$ has been used in the interaction term in Eq. (6.10), and the polarization vectors $\hat{\underset{\sim}{\varepsilon}}$ and $\hat{\underset{\sim}{\varepsilon}}'$

correspond to polarizations α and α' , respectively. The matrix elements in Eq. (7.4) are defined in Eq. (2.13). It is assumed in Eq. (7.4) that the interaction has been switched on at time -T in the remote past. At the end of the calculation the limit as $-T \to -\infty$ will be taken. The quantity $+i0$ in the denominator is interpreted as an imaginary infinitesimal whose imaginary part goes to zero through positive values, and is required for the convergence of the integrals. The total energy change in the process of scattering from the initial state with energy $\varepsilon_n + \hbar\omega$ to the final state with energy $\varepsilon_m + \hbar\omega'$ is

$$\Delta E = \varepsilon_m + \hbar\omega' - \varepsilon_n - \hbar\omega \quad , \tag{7.6}$$

which turns out to be zero when the limit $-T \to -\infty$ is taken.

The probability amplitude $c_m^{(2b)}(t)$ corresponding to Fig. 1 (b) is similar to Eq. (7.4). The electric field amplitudes $\underset{\sim}{E}_o$ and $\underset{\sim}{E}'_o$ are interchanged, since the photon with wave vector $\underset{\sim}{k}'$ and polarization α' is now emitted first. The energy ε_j in the denominator of Eq. (7.4) is replaced by $\varepsilon_j + \hbar\omega + \hbar\omega'$, since there are now two photons in the intermediate state.

To calculate the transition probability per unit time, and then the cross section, it is extremely important that the amplitudes c_m are true gauge-invariant probability amplitudes. We have established this fact in Sec. V. The probability per unit

time w_{fi} of making a transition from the initial state i to the final state f is

$$w_{fi} = \lim_{T\to\infty} \left| c_m^{(2a)}(t) + c_m^{(2b)}(t) \right|^2 / (t + T), \tag{7.7}$$

in second order of perturbation theory, since the numerator is the probability of making a transition during the time t + T. The only part of $c_m^{(2a)}$ and $c_m^{(2b)}$ that depends on the time is the integral in Eq. (7.4), so that the limit in Eq. (7.7) gives

$$\lim_{T\to\infty} \left| \int_{-T}^{'} dt_1 \exp(i\Delta E t_1/\hbar) \right|^2 (t + T)^{-1}$$

$$= 2\pi\hbar\, \delta(\Delta E) \tag{7.8}$$

The δ function whose argument is the change in energy for the scattering process in Eq. (7.6) insures that energy is conserved for the overall scattering process. In contrast, the energy of the intermediate state j in Eq. (7.4) is not equal to the total energy entering

$$\varepsilon_j = \varepsilon_n + \hbar\omega\ , \tag{7.9}$$

since the denominator would then vanish. The state j is thus spoken of as being a virtual state, and for these states the resonance condition in Eq. (2.15) is not valid. Thus in this

case the $-q\underset{\sim}{E}\cdot\underset{\sim}{r}$ perturbation is not equivalent to the $\underset{\sim}{A}\cdot\underset{\sim}{p}$ term of the conventional perturbation theory in Eq. (2.2).

The differential scattering cross section is defined as

$$d\sigma = \sum_{f} w_{fi}/\text{flux} \quad , \tag{7.10}$$

where flux $= c/V$ for a single incoming photon in the volume V, and the sum is over all final states in which a photon is scattered into an element of solid angle $d\Omega$ in the direction of $\underset{\sim}{k}'$. If Eq. (7.4) for $c_m^{(2a)}(t)$ and the similar expression for $c_m^{(2b)}(t)$ are substituted into Eq. (7.7), and Eq. (7.8) is used, the result from Eq. (7.10) for the cross section is

$$d\sigma/d\Omega = r_o^2\, m^2\, \omega(\omega')^2 \hbar^{-2} \left| \sum_j \left(\frac{(\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{r})_{mj}(\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{r})_{jn}}{\omega_{jn} - \omega} + \frac{(\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{r})_{mj}(\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{r})_{jn}}{\omega_{jn} + \omega'} \right) \right|^2 , \tag{7.11}$$

where $r_o = e^2/4\pi mc^2$ is the classical electron radius and ω_{jn} is defined as in Eq. (2.12). Equation (7.11) is the Kramers-Heisenberg (KH) formula for the scattering of a photon by an atomic electron. The two terms within the absolute value bars correspond to the two diagrams of Fig. 1 (a) and (b) with interaction vertices $e\underset{\sim}{E}(0,t)\cdot\underset{\sim}{r}$. This form with the electric dipole

matrix elements is the same form as originally obtained by Kramers and Heisenberg.[20] A derivation along the lines given here is given by Power.[45]

On the other hand, most derivations[42] of the KH formula use the conventional perturbation theory based on the Hamiltonian in Eq. (2.2). The $(e^2/2mc^2)A^2$ term which is represented by the Feynman diagram in Fig. 1 (c), must also be included as a perturbation. The vertices in Fig. 1 (a) and (b) are given in this case by $(e/mc)\underset{\sim}{A}\cdot\underset{\sim}{p}$. A calculation similar to the above gives the cross section[42]

$$d\sigma/d\Omega = r_o^2\,(\omega'/\omega)\left|\delta_{mn}\,\hat{\underset{\sim}{\varepsilon}}'\cdot\hat{\underset{\sim}{\varepsilon}} - (m\hbar)^{-1}\sum_j\left[\frac{(\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{p})_{mj}\,(\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{p})_{jn}}{\omega_{jn}-\omega} + \frac{(\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{p})_{mj}\,(\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{p})_{jn}}{\omega_{jn}+\omega'}\right]\right|^2 . \tag{7.12}$$

The first term within the absolute value bars is the contribution from the A^2 term in Fig. 1 (c), while the next two terms come from Fig. 1 (a) and (b), respectively, with the $\underset{\sim}{A}\cdot\underset{\sim}{p}$ vertex. It is not obvious that Eqs. (7.11) and (7.12) are the same,[46] but Dirac[47] and Kaym[48] have shown that they are. For the sake of completeness the derivation is given in Appendix D using our notation.

It is at first surprising that the gauge dependent conventional perturbation theory would give the same result as the

gauge independent electric dipole interaction. As shown in Appendix A the amplitudes a_n calculated in the conventional perturbation theory are gauge dependent. Therefore, the transition probability calculated from Eq. (7.7) on which the cross section in Eq. (7.10) is based would be incorrect. The resolution of this problem is obtained in Appendix E, where it is shown that the difference between the gauge independent probability amplitudes c_n and the conventional amplitudes a_n is of order e, the charge on the electron. Therefore the discrepancy between the correct electric dipole interaction and the incorrect conventional perturbation theory would show up only in higher-order perturbation theory.

VIII. CONCLUSION

In this paper gauge transformations in the quantum theory of the interaction of radiation and matter have been studied. The special role of the electric dipole interaction $-q\underset{\sim}{E}(0,t)\cdot\underset{\sim}{r}$ in problems in which there is a negligible magnetic field and the wavelength of the radiation is large compared to the size of the system has been emphasized. Calculations involving the electromagnetic field should always be done in a gauge invariant way. In order to use an unperturbed Hamiltonian which does not couple the matter to the radiation, it is necessary to use the electric dipole interaction. Otherwise, the coefficients in the expansion of the wave function in terms of the unperturbed states would be gauge dependent, and cannot be interpreted as probability amplitudes.

The common impression that the expansion coefficients for the wave function can always be interpreted as probability amplitudes is erroneous, since the probability that the system be in a given energy eigenstate must be gauge invariant. Therefore the conventional perturbation theory based on an unperturbed Hamiltonian for the matter uncoupled to the radiation field and using the $\underset{\sim}{A}\cdot\underset{\sim}{p}$ and A^2 interaction terms as perturbations is incorrect, since the expansion coefficients are gauge dependent.

The electric dipole interaction was illustrated by calculating the cross section for the scattering of a photon by an atomic electron. The cross section obtained in this way is equivalent to the Kramers-Heisenberg formula calculated from the conventional perturbation theory using the potentials. The reason for the agreement is that only terms to second order in the fine structure constant are included. In higher orders a discrepancy would appear between the two approaches, and the correct result would be the one in which the electric dipole interaction is used.

The gauge invariant formulation of the perturbation theory discussed in Sec. V is general, and can be extended to problems in which the magnetic field can not be neglected. A gauge invariant formulation in terms of both electric and magnetic multipoles can be given.[17,30]

APPENDIX A: Conventional Perturbation Theory

The customary procedure in solving the time-dependent Schrödinger equation is to use the Hamiltonian in Eq. (2.2) with H_o, defined in Eq. (2.3), as the unperturbed Hamiltonian and the remainder in Eq. (2.2) as the perturbation. The unperturbed Hamiltonian satisfies the Schrödinger equation in Eq. (2.10). The wave function ψ in the Schrödinger equation in Eq. (3.1) can be expanded in terms of $\{\phi_n\}$ as

$$\psi = \sum_n a_n \phi_n \tag{A1}$$

If Eq. (A1) is substituted into the Schrödinger equation in Eq. (3.1), the resulting equation for the coefficients is

$$i\hbar \dot{a}_n = e_n a_n + \sum_m \langle \phi_n | \, | (-q/2mc)(\underset{\sim}{A}\cdot\underset{\sim}{p} + \underset{\sim}{p}\cdot\underset{\sim}{A}) + (q^2/2mc^2)A^2 + qA_o | \phi_m \rangle a_m \quad . \tag{A2}$$

The transitions between states are thus induced by the gauge dependent potentials in Eq. (A2).

If a gauge transformation of the first kind in Eq. (3.2) is made on the wave function, and Eqs. (3.4) and (3.5) are used, the new Schrödinger equation obtained is given in Eq. (3.3). The wave function ψ' can be expanded in terms of

the eigenfunctions in Eq. (2.10) as

$$\psi' = \sum_n a'_n \phi_n \quad . \tag{A3}$$

When Eq. (A3) is substituted into the Schrodinger equation in Eq. (3.3), the equation for the coefficients a'_n obtained is

$$i\hbar\dot{a}'_n = e_n a'_n + \sum_m \langle\phi_n| |(-q/2mc)(\underset{\sim}{A}'\cdot\underset{\sim}{p} + \underset{\sim}{p}\cdot\underset{\sim}{A}') + (q^2/2mc^2)(\underset{\sim}{A}')^2 + qA'_o |\phi_m\rangle a'_m \quad . \tag{A4}$$

The amplitude a'_n is in general not the same as a_n. The question then arises as to what is the correct probability amplitude of finding the system in the state ϕ_n with energy e_n. This question was answered by Lamb[19] when he said that on the basis of experiment the correct gauge to use is the one for which $\underset{\sim}{A} = 0$ as given in Section VI.

That the coefficients a_n and a'_n are not in general the same can be seen by substituting Eqs. (A1) and (A3) into Eq. (3.2) which gives

$$a'_n = \sum_n \langle\phi_n|\exp(iq\Lambda/\hbar c)\phi_m\rangle a_m \quad . \tag{A5}$$

Since $\Lambda = \Lambda(\underset{\sim}{r},t)$, the coefficients in Eq. (A5) will in general be complex numbers not equal to δ_{nm}.

APPENDIX B: Matrix Elements of the Power Operator

In this appendix Eq. (5.7) is derived following the method of Yang.[17] If Eq. (5.1) is differentiated with respect to time, we obtain

$$\dot{H}(\underset{\sim}{A},0)\,\psi_m + H(\underset{\sim}{A},0)\,\dot{\psi}_m = \dot{\varepsilon}_m\psi_m + \varepsilon_m\dot{\psi}_m \ . \tag{B1}$$

If the inner product of Eq. (B1) is taken with respect to ψ_n, the result is

$$\dot{\varepsilon}_m\delta_{nm} + (\varepsilon_m-\varepsilon_n)\langle\psi_n|\dot{\psi}_m\rangle = \langle\psi_n|\dot{H}(\underset{\sim}{A},0)\,\psi_m\rangle \tag{B2}$$

The time derivative of $H(\underset{\sim}{A},0)$ in Eq. (4.9) is

$$\dot{H}(\underset{\sim}{A},0) = (q/2)\left[(\underset{\sim}{E} + \nabla A_o)\cdot\underset{\sim}{v} + \underset{\sim}{v}\cdot(\underset{\sim}{E}+\nabla A_o)\right] \ , \tag{B3}$$

when Eqs. (3.7) and (4.12) are used. However, the commutator of $H(\underset{\sim}{A},0)$ with $q\,A_o$ is

$$[H(\underset{\sim}{A},0),qA_o] = (-i\hbar q/2)\left[\underset{\sim}{v}\cdot\nabla A_o + \nabla A_o\cdot\underset{\sim}{v}\right] \tag{B4}$$

Therefore Eq. (B3) can be written as

$$\dot{H}(\underset{\sim}{A},0) = P + (i\hbar)^{-1}[qA_o,H(\underset{\sim}{A},0)] \ , \tag{B5}$$

where P is the power operator defined in Eq. (4.15).

Equation (B2) then becomes

$$\dot{\varepsilon}_m \delta_{nm} + (i\hbar)^{-1}(\varepsilon_n - \varepsilon_m)\langle\psi_n|(qA_o - i\hbar\partial/\partial t)\psi_m\rangle$$

$$= \langle\psi_n|P\psi_m\rangle \quad . \tag{B6}$$

For $n \neq m$ Eq. (5.7) is obtained, while for $n = m$, we have a special case of Eq. (4.14).

AD-A048 194 NORTH TEXAS STATE UNIV DENTON DEPT OF PHYSICS F/G 20/12
OPTICALLY INDUCED HOT ELECTRON EFFECTS IN SEMICONDUCTORS.(U)
AUG 77 A L SMIRL N00014-76-C-1077
UNCLASSIFIED NL

3 OF 3
AD A048194

END
DATE FILMED
1 -78
DDC

APPENDIX C: Power Operator and Electric Dipole Energy

In this appendix Eq. (5.11) is derived. A straightforward calculation shows that the commutator of $\underset{\sim}{r}$ and $H(\underset{\sim}{A},0)$ is

$$[r,H(\underset{\sim}{A},0)] = i\hbar\underset{\sim}{v} \quad , \tag{C1}$$

where the velocity operator $\underset{\sim}{v}$ is defined in Eq. (4.12). If Eq. (C1) is substituted into the matrix element of the power operator in Eq. (4.15), the result is

$$\begin{aligned}\langle\psi_n|P\psi_m\rangle = (q/2i\hbar)\{&(\varepsilon_m-\varepsilon_n)\langle\psi_n|\underset{\sim}{E}\cdot r\psi_m\rangle \\ &-\langle\psi_n|\left(E_iH(\underset{\sim}{A},0)x_i - x_iH(\underset{\sim}{A},0)E_i\right)\psi_m\rangle\}\end{aligned} \tag{C2}$$

where Eq. (5.1) has been used.

If the commutator of E_i and $H(\underset{\sim}{A},0)$ is calculated the result is

$$[E_i,H(\underset{\sim}{A},0)] = (i\hbar/2)\left(v_j(\partial E_i/\partial x_j) + (\partial E_i/\partial x_j)v_j\right) \tag{C3}$$

for $i = 1,2,3$, where $v_j (j = 1,2,3)$ are the components of the velocity operator v in Eq. (4.12). When Eq. (C3) is substituted into the last term in Eq. (C2), the result is

$$\langle\psi_n|\left(E_i H(\underset{\sim}{A},0)x_i - x_i H(\underset{\sim}{A},0)E_i\right)\psi_m\rangle$$

$$= (\varepsilon_n-\varepsilon_m)\langle\psi_n|\underset{\sim}{E}\cdot\underset{\sim}{r}\psi_m\rangle$$

$$+ (i\hbar/2)\langle\psi_n|\{\{v_j,\partial E_i/\partial x_j\},x_i\}\psi_m\rangle\ , \qquad \text{(C4)}$$

where the curly brackets are anticommutator brackets defined in Eq. (5.12). If Eq. (C4) is substituted into Eq. (C2), we obtain Eq. (5.11).

APPENDIX D: Equivalence of the Two Forms of the Kramers-Heisenberg Formula

The two forms of the KH formula given in Eqs. (7.11) and (7.12) have been derived using the $e\underset{\sim}{E}\cdot\underset{\sim}{r}$ perturbation in Eq. (2.5) and the $(e/mc)\underset{\sim}{A}\cdot\underset{\sim}{p} + (e^2/2mc^2)A^2$ perturbation in Eq. (2.2), respectively. Their equivalence has been shown by Dirac[47] and Baym,[48] but for the sake of completeness the derivation is given here using our notation.

The commutation relation between the components of $\underset{\sim}{r}$ and $\underset{\sim}{p}$ in Eq. (2.8) implies that the Hamiltonian H_o in Eq. (2.3) and $\underset{\sim}{r}$ have the commutator given in Eq. (2.9). These two commutation relations can be used to show that the first term on the right hand side of Eq. (7.12) is

$$\hat{\underset{\sim}{\varepsilon}}'\cdot\hat{\underset{\sim}{\varepsilon}}\delta_{mn} = (m\hbar)^{-1} \sum_j \{\omega_{jn}^{-1} (\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{p})_{mj}(\varepsilon\cdot\underset{\sim}{p})_{jn} - \omega_{mj}^{-1} (\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{p})_{mj}(\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{p})_{jn}\} \qquad (D1)$$

If Eq. (D1) is substituted into Eq. (7.12) and Eq. (2.9) is used, the cross section is

$$d\sigma/d\Omega = (r_o^2\, m^2\, \omega'/\omega\hbar^2) \times \left| \sum_j \left\{ (\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{r})_{mj}(\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{r})_{jn}\omega_{mj}\omega_{jn}\left(\omega_{jn}^{-1} - (\omega_{jn}-\omega)^{-1}\right) + (\hat{\underset{\sim}{\varepsilon}}\cdot\underset{\sim}{r})_{mj}(\hat{\underset{\sim}{\varepsilon}}'\cdot\underset{\sim}{r})_{jn}\omega_{mj}\omega_{jn}\left(-\omega_{mj}^{-1} - (\omega_{jn}+\omega')^{-1}\right)\right\}\right|^2 . \qquad (D2)$$

Using Eq. (2.12) and the energy conservation implied by substituting Eq. (7.6) into Eq. (7.8) we can prove that

$$\omega_{mj}\omega_{jn}\left(\omega_{jn}^{-1} - (\omega_{jn}-\omega)^{-1}\right) = \omega\omega'(\omega_{jn}-\omega)^{-1} + \omega \qquad \text{(D3)}$$

and

$$\omega_{mj}\omega_{jn}\left(-\omega_{mj}^{-1} - (\omega_{jn}+\omega')^{-1}\right) = \omega\omega'(\omega_{jn}+\omega')^{-1} - \omega \ . \qquad \text{(D4)}$$

Equations (D3) and (D4) can be substituted into Eq. (D2), which gives Eq. (7.11), except for an extra term. That extra term vanishes,

$$\sum_j \left[(\hat{\epsilon}'\cdot\underset{\sim}{r})_{mj}(\hat{\epsilon}\cdot\underset{\sim}{r})_{jn} - (\hat{\epsilon}\cdot\underset{\sim}{r})_{mj}(\hat{\epsilon}'\cdot\underset{\sim}{r})_{jn}\right] = 0, \qquad \text{(D5)}$$

since the atomic states are complete, and the coordinates commute. Therefore, we have proved the equivalence of the two forms of the KH formula in Eqs. (7.11) and (7.12).

APPENDIX E: Relation Between Gauge Dependent and Independent Amplitudes

In this appendix it is shown why the conventional perturbation theory discussed in Sec. II and Appendix A gives the same KH cross section in Eq. (7.12) as the KH formula calculated in Eq. (7.11). The equality is established in Appendix D. It is shown that the difference between the coefficients is proportional to the square root of the fine structure constant $e^2/\hbar c$, so that the discrepancy between perturbation theory with the electric dipole interaction and the conventional perturbation theory will show up only in higher orders.

The gauge function $\Lambda(\underset{\sim}{r},t)$ relating the electric dipole interaction in Eq. (2.5) with the Hamiltonian in Eqs. (2.1) or (2.2) is

$$\Lambda(\underset{\sim}{r},t) = -\underset{\sim}{A}(0,t)\cdot\underset{\sim}{r} + \Lambda(0,t), \tag{E1}$$

from Eq. (6.2). The coefficients a_n' in Eqs. (A3) and (A5) are, in this case, the coefficients c_n of Eq. (5.5). Thus the gauge invariant coefficients c_n are related to the gauge dependent coefficients a_n of the conventional perturbation theory by

$$c_n = \exp\left(iq\Lambda(0,t)/\hbar c\right)\sum_m <\phi_n|\exp\left(-i(q/\hbar c)\underset{\sim}{A}(0,t)\cdot\underset{\sim}{r}\right)\phi_m>a_m . \tag{E2}$$

from Eq. (A5). If the exponential in Eq. (E2) is expanded, we obtain

$$c_n = a_n - i(q/\hbar c) \sum_m \underset{\sim}{A}(0,t)\cdot(\underset{\sim}{r})_{nm} a_m + \cdots \quad . \tag{E3}$$

Thus the difference between the gauge independent probability amplitude c_n and the gauge dependent expansion coefficient a_n is of order $q = -e$, the charge on the electron. Therefore, the conventional perturbation theory gives the same result for the KH formula in Eq. (7.12) as the gauge independent form in Eq. (7.11) to second order in the fine structure constant $\alpha = e^2/\hbar c$.

REFERENCES AND NOTES

1. W. Heisenberg, Zeit. f. Phys. 33, 879 (1925).
   E. Schrödinger, Ann. Phys. (Leipzig) 81, 109 (1926).
2. L. I. Schiff, Quantum Mechanics (McGraw-Hill Book Co., New York, 1968) Third ed., pp. 399, 423.
3. E. Merzbacher, Quantum Mechanics (John Wiley and Sons, Inc., New York, 1970) Second ed., p. 170.
4. K. Gottfried, Quantum Mechanics (W. A. Benjamin, Inc., New York, 1966), Vol. I, p.75.
5. R. H. Dicke and J. P. Wittke, Introduction to Quantum Mechanics (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1960), p. 263.
6. See, e.g., Ref. 2, pp. 400-401, Ref. 3, pp. 458-459; Ref. 4, p. 465.
7. M. Mizushima, Quantum Mechanics of Atomic Spectra and Atomic Structure (W. A. Benjamin, Inc., New York, 1970), pp. 72-79.
8. D. Bohm, Quantum Theory (Prentice-Hall, Inc., New York, 1951), p. 417.
9. H. A. Bethe and R. Jackiw, Intermediate Quantum Mechanics (W. A. Benjamin, Inc., New York, 1968) Second ed., pp. 191, 207.
10. O. Svelto, Principles of Lasers (Plenum Press, New York, 1976), p. 27.
11. L. Pauling and E. B. Wilson, Jr., Introduction to Quantum Mechanics (McGraw-Hill Book Co., New York, 1935), pp. 302-306.

12. M. Sargent, III, M. O. Scully, and W. E. Lamb, Jr., Laser Physics (Addison-Wesley Publishing Co., Reading, Mass., 1974), pp. 15-16.

13. R. H. Pantell and H. E. Puthoff, Fundamentals of Quantum Electronics (John Wiley and Sons, Inc., New York, 1969), pp. 20-26.

14. G. Baym, Lectures On Quantum Mechanics (W. A. Benjamin, New York, 1969), pp. 297-9.

15. W. H. Louisell, Quantum Statistical Properties of Radiation (John Wiley and Sons, New York, 1973), pp. 514-517.

16. R. Loudon, The Quantum Theory of Light (Clarendon Press, Oxford, 1973), pp. 164-173.

17. K.-H. Yang, Ann. Phys., (N.Y.) 101, 62 (1976).

18. The radiation may interact with any type of matter - nucleus, atom, molecule, or solid, but for the sake of definiteness atom is used when dealing with the material system. The material system is, however, treated nonrelativistically here, although a relativistic treatment can easily be given.

19. W. E. Lamb, Jr., Phys. Rev. 85, 259 (1952).

20. H. A. Kramers and W. Heisenberg, Zeit. f. Phys. 31, 681 (1925).

21. V. Fock, Zeit. f. Phys. 39, 226 (1927).

22. H. Weyl, Gruppentheorie und Quantenmechanik (S. Hirzel, Leipzig, 1931), Second. ed., The Theory of Groups and Quantum Mechanics (Methuen & Co., London, 1931), (reprinted by Dover Publications, Inc., N.Y., 1950), pp. 100, 214-220.

23. See, e.g., E. S. Abers and B. W. Lee, Phys. Repts. 9C, 1-141 (1973).

24. W. Pauli, Rev. Mod. Phys. 13, 203 (1941).

25. The literature is much too vast to be comprehensive, but the following references are typical.

J. Schwinger, Phys. Rev. 82, 664 (1951).

F. J. Belinfante, Phys. Rev. 84, 648, 546, 541 (1951).

P. A. M. Dirac, Ann. Inst. Henri Poincare (Paris) 13, 1 (1952).

I. Goldberg, Phys. Rev. 112, 1361 (1958).

I. Goldberg and E. Marx, Nuovo Cimento B 57, 485 (1968).

26. This is not the place to go into detail on the technical difficulties, but the following references are typical.

F. Strocchi, Phys. Rev. D 2, 2334 (1970).

I. Bialynicki-Birula, Phys Rev. D 2, 2877 (1970).

27. Y. Aharonov and D. Bohm, Phys. Rev. 115, 485 (1959); ibid. 123, 1511 (1961); ibid., 125, 2192 (1962); ibid., 130, 1625 (1963).

28. E. A. Power and S. Zienau, Phil. Trans. Roy. Soc. (Ser. A) 251, 427 (1959).

29. See, e.g., J. Fiutak, Can. J. Phys. 41, 12 (1963).

P. W. Atkins and R. G. Wooley, Proc. Roy. Soc (London) A319, 549 (1970).

J. Savolainen and S. Stenholm, Am. J. Phys. 40, 667 (1972).

30. R. G. Woolley, Adv. Chem. Phys. 33, 153 (1975).

31. A number of quantum mechanics texts (see Refs. 6-8) deal only with radiation problems for which the A^2 term can be neglected, but this gauge dependent approximation should not be made.

32. J. J. Sakurai, Advanced Quantum Mechanics (Addison-Wesley Publishing Co., Reading, Mass., 1967), pp. 41-42.

33. Reference 19. On p. 268 Lamb writes, "If, however, the perturbation is written in terms of a vector potential $\underset{\sim}{A}$ on the rf field as $-(e/m)\underset{\sim}{A}\cdot\underset{\sim}{p}$ instead of $e\underset{\sim}{E}\cdot\underset{\sim}{r}$ an additional factor of $(\omega/\nu)^2$ appears . . . . The usual interpretation of probability amplitudes is valid only in the former gauge [$e\underset{\sim}{E}\cdot\underset{\sim}{r}$], and no factor of $(\omega/\nu)^2$ actually occurs." Then he thanks L. H. Thomas for discussions on these questions. The factor (ω/ν) referred to by Lamb is the factor (ω_{nm}/ω) in our Eq. (2.14).

34. See, e.g., M. Goeppert-Mayer, Ann. Physik 9, 273 (1931); V. F. Weisskopf and E. Wigner, Zeit. f. Phys. 63, 54 (1930).

35. See, e.g., Ref. 8, pp. 357-8; Ref. 14, pp. 264-5; and Ref. 22.

36. H. A. Kramers, Quantum Mechanics, North Holland Publishing Co., Amsterdam, 1957), (reprinted by Dover Publications, Inc., New York, 1964), pp. 268-270.

37. W. Pauli, in Encyclopedia of Physics, S. Flügge, ed. (Springer-Verlag, Berlin, 1958), Vol. 5, pt. 1, pp. 29-30.

38. See, e.g., J. D. Jackson, Classical Electrodynamics (John Wiley and Sons, Inc., New York, 1975), Second ed., pp. 219-220.

39. This fact was pointed out by P. A. M. Dirac, Can. J. Phys. 33, 650 (1955).

40. See, e.g., Ref. 2, pp. 169-170; Ref. 3, pp. 165, 337.

41. See, e.g., Ref. 2, pp. 178-179; Ref. 37, pp. 24-29.

42. See, e.g., Ref. 32, pp. 47-50.

43. The KH formula was obtained from wave mechanics by E. Schrödinger (Ref. 1) without quantizing the electromagnetic field.

44. See, e.g., Ref. 32, p. 29.

45. E. A. Power, Introductory Quantum Electrodynamics (Longmans, London, 1964), pp. 107-8.

46. There is some confusion regarding the equivalence of these two forms. For example, W. Heitler [The Quantum Theory of Radiation (Oxford University Press, London, 1954), Third ed., p. 192] gives credit to I. Waller [Zeit. f. Phys. 31, 681 (1925)] for discovering the $\hat{\underset{\sim}{\varepsilon}}' \cdot \hat{\underset{\sim}{\varepsilon}}$ term in Eq. (7.12). Waller used what is now the conventional perturbation theory in terms of the vector potential $\underset{\sim}{A}$, keeping the A^2 term in addition to the $\underset{\sim}{A} \cdot \underset{\sim}{p}$ term, and derived Eq. (7.12). On the other hand, Kramers and Heisenberg (Ref. 20) used the electric dipole interaction and derived Eq. (7.11). However, Heitler implies that Kramers and Heisenberg derived Eq. (7.12) without the $\hat{\underset{\sim}{\varepsilon}}' \cdot \hat{\underset{\sim}{\varepsilon}}$ term.

47. P. A. M. Dirac, The Principles of Quantum Mechanics (Oxford University Press, London, 1958) Fourth revised ed., pp. 244-248.

48. Ref. 14, pp. 289-96.

LIST OF FIGURES

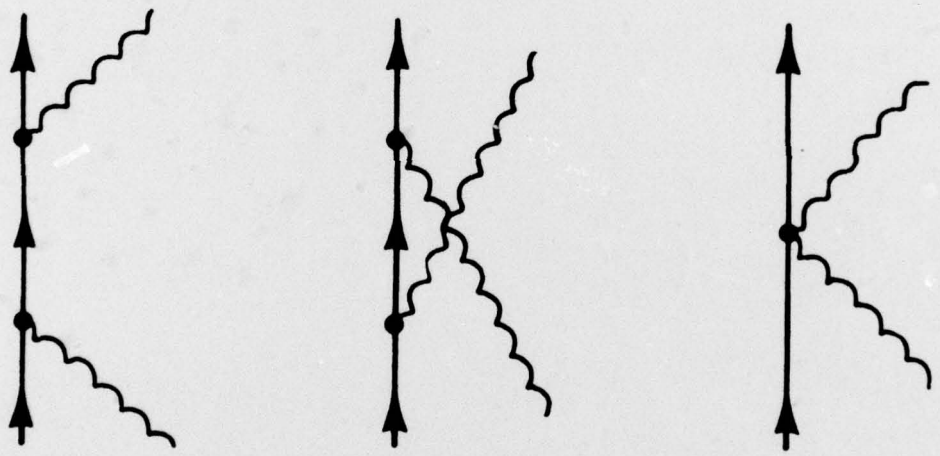


FIGURE 1