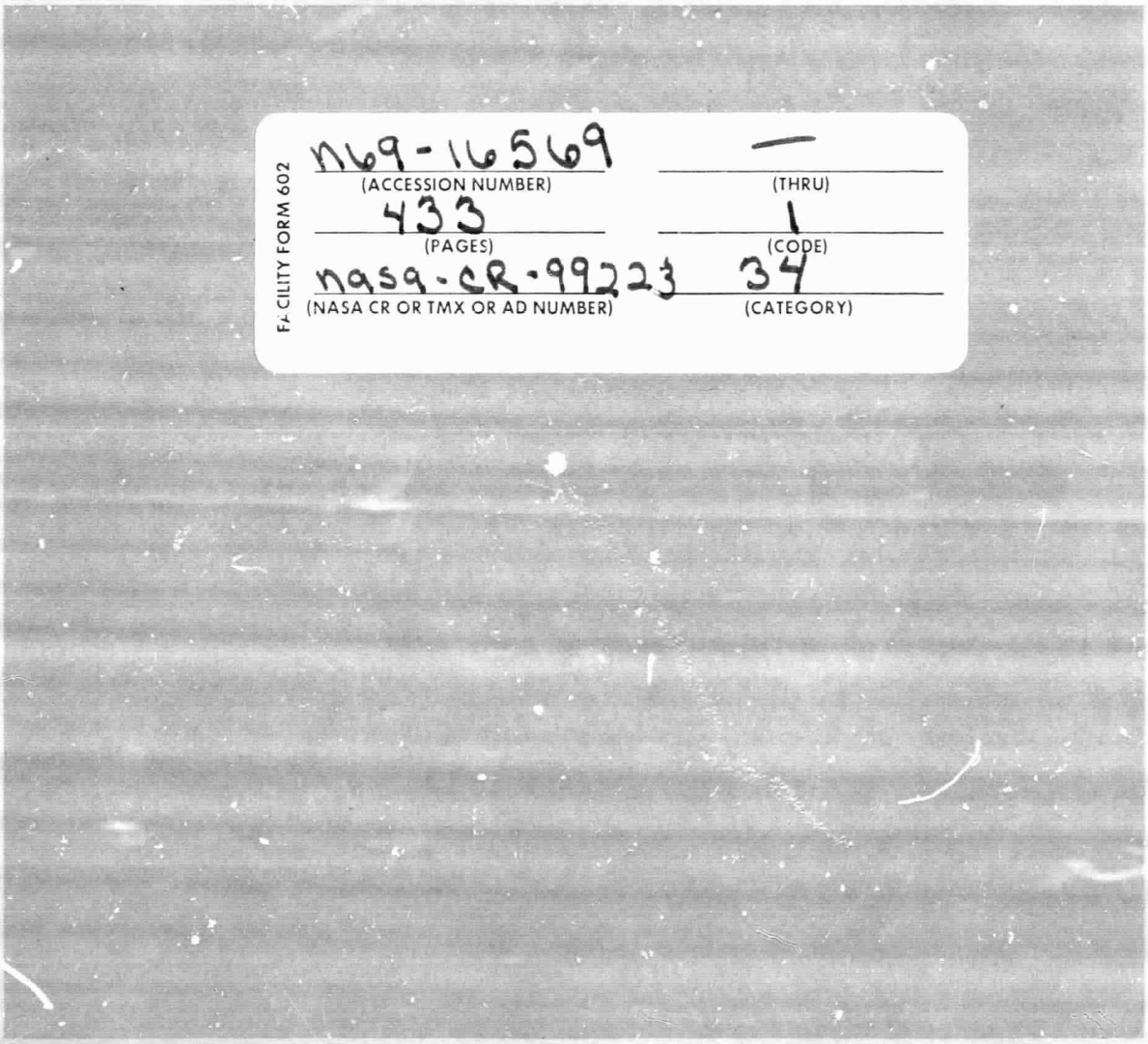
SEPTEMBER 30, 1968

CSL COORDINATED SCIENCE LABORATORY

PROGRESS REPORT FOR MARCH THROUGH AUGUST, 1968



UNIVERSITY OF ILLINOIS – URBANA, ILLINOIS

The research reported in this document was made possible through support extended the Coordinated Science Laboratory, University of Illinois, by the Joint Services Electronics Program (U. S. Army Electronics Command and U. S. Army Research Office, Office of Naval Research and the Air Force Office of Scientific Research) under Contract Number DAAB-07-67-C-0199.

Portions of this work were also supported by:

National Aeronautics and Space Administration

Grant NsG 228-62

National Science Foundation

Grant NSF GK-1663, Grant NSF GK-2339

Advanced Research Projects Agency

Through U. S. Army Contracts DAAK-02-67-C-0546

and DAAB-07-67-C-0199

Air Force Office of Scientific Research

Grant AFOSR 931-67
Grant AFOSR 68-1579

Office of Naval Research

Contracts N00014-67-A-0305-0001, NONR Symposium

U. S. Office of Education

Contract OE C-1-7-071213-4557

Ford Motor Company

Owens-Illinois, Incorporated

Physical Electronics Affiliates Program

College of Engineering, University of Illinois

University of Illinois

as acknowledged in footnotes in the text

(OVERLEAF BLANK)

SUMMARY OF PROGRESS REPORT FOR MARCH THROUGH AUGUST 1968

1. Surface Physics

Measurements of the emission of ions from polycrystalline tungsten due to electron bombardment have been initiated. Relatively complete studies of the ion-energy distributions, adsorption and desorption characteristics, ion-desorption cross section, and neutral-desorption cross section have been made for the system of oxygen adsorbed on polycrystalline tungsten. Initial measurements have been made for hydrogen, carbon monoxide, and nitrogen on polycrystalline tungsten have been made. The oxygen measurements are summarized. Polycrystalline platinum has been investigated using the high-resolution spectrometer. The results indicate that the various cleaning techniques did not yield clean surfaces and that the adsorption characteristics were substantially influenced because of this. These measurements are discussed. Initial diffraction, elastic-scattering, and inelastic-scattering measurements have been made utilizing the angular energy distribution apparatus. These measurements show that the types of results which were expected are in fact realizable. Problems in producing a clean, well-ordered surface for this experiment are discussed.

2. Infrared Converter and Applied Physics

Development of the solid-state infrared converter has continued to emphasize component materials. Data on resistivity and spectral properties of CdS films formed both by vacuum deposition and by transport are reported. Films of the photoconductors PbTe and PbS have been made by vacuum evaporation, and rough measurements of detectivities are given.

Both experimental and theoretical attention has been given to obtaining information about tunneling barriers. Some information is given about discrete (nonplanar) light-emitting diodes, and their incorporation into a simple infrared converter. In electrical size-effect research, measurements on gold films have continued with more sophisticated equipment. A theoretical model pertinent to size effects in semimetals was examined and discarded.

3. Plasma Physics

Unstable oscillations at frequencies below the plasma frequency are observed at the plasma boundary during the interaction of an electron beam with a plasma. These oscillations are interpreted as the eigenmodes of the inhomogeneous plasma boundary, driven into instability by the beam. It is shown that electron beams suitable for experiments on weak turbulence can be produced by passing the beams through aluminum foils nearly one micron thick. Measurements of the beam distribution function and the transmission as a function of foil thickness are presented. Collisionless cesium and potassium plasmas are investigated in a single-ended Q machine. A nonlinear interaction of two different drift waves is observed and discussed.

4. Rarefied Gas Dynamics

Monte-Carlo solutions of the nonlinear Boltzmann equation for shock waves have been obtained for Mach numbers from 1.05 to 10 and have been compared with 6-moment and Navier-Stokes models where appropriate. Monte-Carlo solutions of this equation for heat transfer in a rarefied gas

have been extended to cover Knudsen numbers from 0.1 to 100 and temperature ratio from 0.1 to 0.9.

5. High-Voltage Breakdown

A statistical study of surface parameters associated with pre-breakdown field emission and electrical breakdown sheds some light on the nature of the cathode whiskers responsible for initiation of a breakdown. There is a distribution in heights and a correlated distribution in critical breakdown field at the whisker tips, in general agreement with theory. Gas-conditioning experiments indicate that during conditioning many new inhomogeneous whiskers are formed by sputtering, and that, although the breakdown voltage is increased, the critical field at the whisker tips is reduced. Work with the cylindrical field-emission microscope shows that to form new whiskers on the tungsten center wire at room temperature, a nearby anode must be present, probably to supply gas ions for the sputtering process. New crystalline whiskers are also formed at a wire temperature of 750°C. Several computer programs have been developed. These enable rapid computation of surface parameters and automatic plotting of the Fowler-Nordheim plots, thus making practical the statistical studies reported above.

6. Space and Atmospheric Sciences

The development of optimal orbit and spin-orientation parameters based on a new data-analysis simulation for the relativity satellite is described. The feasibility study for the relativity experiment was successfully completed with the issuance of a final report, although some few investigations have continued, and a proposal to perform the experiment

is in preparation. A feasibility study of the use of a similar satellite for the purpose of measuring gas-surface interaction parameters at orbital speeds is outlined. New equipment and preliminary measurements for a program of interferometric measurements of far-infrared spectra are described.

7. Semiconductor Physics

The influence of defects and impurities upon the properties of silicon and germanium are under investigation. Results are presented of the measurements on the recombination luminescence of irradiated silicon, of the tunneling between metal-semiconductor junctions, and of the electrical properties of p-i-n diodes.

8. Computer

Instrumentation for preparing computer-controlled motion pictures, further development of the language CSL6, and the adaptation of the Cooley-Tukey Fast Fourier Transform algorithm to CSL programs are reported.

9. Plasma Display

A procedure for obtaining variable intensity, a theory for the stability of firing times in the Plasma Display cell, and a new approach to circuit design in which the functions of addressing and decoding are combined in a way that reduces the required number of active elements are reported.

10. Urban Development

An in-depth study of the factors causing the growth and decay of Kankakee, Illinois has been started. During this first phase of the study, data havebeen collected on the location, price and date of real estate

transactions throughout the Kankakee Metropolitan Area. Approximately one-half of the data have been prepared for computer analysis and display.

11. Information Science

This section covers our progress in information science and systems. Summarized are contributions in coding theory and computer reliability, digital and stochastic systems, communications technology, and retrieval systems.

12. Networks

Some results are reported on the application of optimization techniques in the computer-aided design of filters. Two basic approaches were taken, and the advantages and disadvantages of each are discussed. Also, studies are continuing in the stability of nonlinear networks. Some interesting results have been obtained regarding the asymptotic stability of a network in which is embedded a nonlinear capacitor. The characterization of biplanar graphs is also being studied.

13. Control Systems

Investigation of parameter-variation effects (sensitivity) in control-system performance is continuing. Analysis and design of low-sensitivity systems has been investigated, using sensitivity models, singular perturbation techniques, differential game theory, approximate optimization, and minimax design. Simulation studies of two large-scale systems have been carried out.

14. Switching Systems

The NAND-decomposition algorithm has been extended to permit the specification of network constraints such as fan-in, fan-out, and

level limits, and results obtained with and without network constraints are compared. Preliminary results on the relative effectiveness of modular-decomposition approaches and the NAND decomposition followed by packaging are reported. Investigations of methods for the generation of diagnostic tests continue with emphasis on module-level diagnosis. New investigations into the design of diagnosable sequential machines and the diagnosis of cellular arrays have been started. The study of the applicability of ternary logic to speed-independent sequential networks continues.

15. Cognitive Memory

This project is a study of the fundamentals of cognitive processes and its application to information-retrieval systems. Progress reported includes investigations of a general question-answering system, universal data structures, nonlinguistic cognition, associative processors, and mathematics of cognition.

16. Seismic Detection

Additional field tests and computer analyses of vehicular seismic signals have been carried out in an attempt to better understand the relationships between spectral frequency components and specific features of the vehicle.

COORDINATED SCIENCE LABORATORY PERSONNEL

Faculty, Research Associates, and Research Engineers

Compton, W. D.
Director
Alpert, D.
Anner, G. E.
Ash, R. B.
Bitzer, D. L.
Bohmer, H.
Bouknight, W. J.
Brown, R. M.
Carroll, D. E.
Carss, B. W.
Chien, R. T.
Cooper, D. H.
Cruz, J. B., Jr.
Culton, J. W.
Ettinger, S. Y.
Fenves, S. J.
Gooch, J.
Haddad, A. H.
Hicks, B. L.
Hohn, F.
Jackson, E. A.
Jacobs, J. T.
Kelley, K. C.
Kirkwood, B. D.
Knoebel, H. W.
Kokotovic, P.
Krone, H. V.
Lee, D. A.
Lyman, E. M.
Lyman, E. R.
Mayeda, W.
Medanic, J.
Metze, G.
Peacock, R. N.
Perkins, W. R.
Preparata, F. P.
Propst, F. M.
Prothe, W. C.
Assistant to the Director
Raether, M.
Ray, S.
Resch, R. D.
Scott, E. J.
Skaperdas, D.
Slottow, H. G.
Steinrisser, F.
Stifle, J. E.
Trick, T. N.
Trogdon, R.
Tulumello, A.
Vaughn, W. H.
Von Foerster, H.
Voss, J. R.
Voth, B. W.
Wax, N.
Yen, S. M.

Research Assistants

Aprille, T. J., Jr.
Arora, B. M.
Bahl, L. R.
Bielowski, H.
Biss, K. O.
Bleha, W. P., Jr.
Bollinger, L. D.
Brachhausen, E. H.
Brown, W. F.
Burkstrand, J. M.
Chambers, R. S.
Chang, J. S.
Chen, C. I.
Chow, J. C.
Chu, W. H.
Cooper, T. L.
Cullen, D. E.
Cummings, J. W.
Davidson, E. S.
Depp, S. W.
Desroches, J. P.
Duran, P. S.
Edwards, D., Jr.
Evans, K. E., Jr.
Fenneman, D. B.
Foster, C. A.
Gardner, M. F.
Goede, W. F.
Hartmann, C. R.
Hatfield, F. J.
Hayes, J. M.
Hecht, R. W.
Heller, J. E.
Holden, R. A.
Hong, S. J.
Hosken, R. W.
Jamskidi, M.
Jansen, J. M.
Jimenez, R. C.
Johnson, E. S.
Johnson, H. R.
Johnson, R. L.
Jones, C. E.
Karr, G. R.
Kisylia, A. P., Jr.
Kraybill, D. M.
Laggan, J. J.
Lombardi, D. J.
Metze, V. C.
Moore, J. K.
Myers, J. L., Jr.
Ng, S. W.
Nicholas, M. J.
Nishijima, M.
Rajput, B. S.
Reilly, B.
Reilly, O. G., Jr.
Robinson, J. E.
Rose, R. E.
Sabino, A. C.
Sakurai, J.
Sannuti, P.
Schein, L. B.
Schermerhorn, J. D.
Schultz, J. A.
Seth, S. C.
Stahl, F. A.
Stefanek, R. G.
Stewart, C. N.
Tao, B.
Toida, S.
Tzeng, K. K.
Wang, B. S.
Weston, P. E.
Wilkie, D. F.
Wojcik, A. S.
Woo, T.

Fellows

Gaddess, T.
Lipovski, G.
Miller, D.
Powell, T.
Roth, K.
Schertz, D.

Administrative Secretary

Schmidt, R. F.

Typists and Stenographers

Anson, J.
Champagne, B.
Hanoka, N.
Kallembach, S.
Kempton, B.
Lahey, M.
Lane, R.
Siler, J.

Chief Clerk

Drews, C.

Duplicating Machine Operator

Potter, R.

Chief Engineering Draftsman

MacFarlane, R.

Draf sman

Harmon, D.

Research Lab. Shop Supervisor

Bandy, L.

Research Engineering Asst.

Burr, J. G.

Laboratory Mechanics

Bales, R.
Beaulin, W.
Bouck, G.
Fults, R.
Merritt, K.

Stores Supervisor

Lofton, C.

Storekeepers

Jordan, F.
McElwain, W.

Glassblower

Lawrence, W.

Photographer

Gladin, R. T.

Electronics Engr. Assts.

Carter, E.
Gardner, O. E.
Hedges, L.
Neff, E.
Vassos, N.

Electronics Technicians

Casale, T.
Coad, D.
Crawford, G.
Deschene, D.
Holy, F.
Johnson, M.
Knoke, J.
Merrifield, F.
Moule, G.
Roberts, G.
Schmidt, W.
Streff, L.
Susedik, A.
Turpin, F.

Physical Sci. Staff Asst.

Thrasher, W.

Student Assistants

Arnold, C.
Baran, J.
Birtcher, R.
Bonacci, J.
Chew, R.
Coates, W.
Dehn, M.
Duchamp, T.
Erbes, D.
Fariss, G.
Gates, M.
Goodman, G.
Hammerton, D.
Hansen, J.
Hoffman, R.
Jones, R.
Kubala, R.
Lewis, R.
Loew, F.
Omundson, J.
Plunk, M.
Reynolds, P.
Roberts, D.
Rundgren, W.
Stupp, G.
Taylor, S.
Tucker, P.
Veach, R.
Walhus, P.
Ward, W.
Zbylut, R.

1. Journal Articles Published or Accepted

R. T. Chien, D. T. Tang, E. S. Barrekette, and A. M. Katcher, "Analysis and Improvement of Photostore Error Rates," Proc. IEEE 56, 805 (May, 1968).

G. R. Karr and S. M. Yen, "Aerodynamic Torque on a Spinning Spherical Satellite with Application to Measurement of Accommodation Coefficients," Astronautica Acta, in press (Mar.-April, 1968).

P. V. Kokotović, J. B. Cruz, Jr., J. E. Heller, and P. Sannuti, "Synthesis of Optimally Sensitive Systems," Proceedings of the IEEE, 56, in press (August 1968).

P. Kokotović, J. Heller, P. Sannuti, "Sensitivity Comparison of Optimal Controls," Int. J. Control, in press (Sept-Oct, 1968).

P. Kokotović and P. Sannuti, "Singular Perturbation Method for Reducing the Model Order in Optimal Control Design," IEEE Trans. on Automatic Control AC-13, in press (August 1968).

A. H. Haddad, "On a Nonlinear Feedback Filtering System," IEEE Trans. on Information Theory, IT-14, 532 (July 1968).

A. H. Haddad, "Optimum Filtering with a Class of Nonlinear Systems," IEEE Trans. on Automatic Control, AC-13, 289 (June 1968).

A. H. Haddad and P. V. Kokotović, "Estimators for Optimally Sensitive Systems," Proc. IEEE, 56, in press (August 1968).

Bruce L. Hicks and Margaret A. Smith, "On the Accuracy of Monte-Carlo Solutions of the Nonlinear Boltzmann Equation," Journal of Computational Physics 3, in press (1968).

H. T. Hsu, T. Kasami, and R. T. Chien, "Error-Correcting Codes for a Compound Channel," IEEE Trans. IT-14, 135.

F. P. Preparata, "Convolutional Transformation and Recovery of Binary Sequences," IEEE Trans. on Computers C-17, 649 (July 1968).

F. P. Preparata, "Weight and Distance Structure of Nordstrom-Robinson Quadratic Code," Information and Control, in press (May 1968).

W. Mayeda, S. L. Hakimi, W. Chen, and N. Deo, "Generation of Complete Trees," IEEE Trans. on Circuit Theory CT-15, in press (June 1968).

R. A. Werner and J. B. Cruz, Jr., "Feedback Control Which Preserves Optimality for Systems with Unknown Parameters," IEEE Trans. on Automatic Control, AC-13, in press (December 1968).

2. Meeting Papers

R. T. Chien and K. Tzeng, "On Iterative Decoding of BCH Codes," Proc. Eleventh Midwest Symposium on Circuit Theory, 599 May, 1968.

D. Chow and R. T. Chien, "Codes from Finite Geometry," Presented at the Second Annual Princeton Conference on Information Sciences and Systems March, 1968.

A. H. Haddad, "On a Class of Step Processes," Proc. 2nd Annual Princeton Conference on Information Sciences and Systems, 169, March, 1968.

B. L. Hicks and S. M. Yen, "Solution of the Nonlinear Boltzmann Equation for Plane Shock Waves," 6th International Symposium on Rarefied Gas Dynamics, MIT, Cambridge, Mass., July 22-26, 1968.

P. Kokotović and P. Sannuti, "Sensitivity Method for Reducing the Model Order in Optimal Control Design," Joint Automatic Control Conference, Ann Arbor, Michigan, June 26-28, 1968.

P. Kokotović and R. Stefanek, "Obtaining Fuel-Optimal Controls for Linear Time-Varying Plants by Newton's Method," 2nd International Conf. on Computing Methods in Optimization Problems, San Remo, Italy, September 1968.

P. Sannuti and P. Kokotović, "An Approximate Design of Optimal Regulators for High-order Linear Plants," IFAC Symposium on System Sensitivity and Adaptivity, Dubrovnik, Yugoslavia, August 26-31, 1968.

R. G. Stefanek and J. B. Cruz, Jr., "Use of Switch Time Sensitivities to Provide Near Optimal Control," IFAC Symposium on System Sensitivity and Adaptivity, Dubrovnik, Yugoslavia, Aug 26-31, 1968.

T. N. Trick, "Stability of Networks Containing Nonlinear Time-Varying Capacitors," 11th Midwest Symposium on Circuit Theory, University of Notre Dame, 352, May, 1968.

T. N. Trick, "Stability of Networks Containing Nonlinear Time-Varying Capacitors," Summer School on Circuit Theory 1968, Institute of Radio Engineering and Electronics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia.

D. F. Wilkie and W. R. Perkins, "Essential Parameters in Sensitivity Analysis," IFAC Symposium System Sensitivity and Adaptivity, Dubrovnik, Yugoslavia, August 26-31, 1968.

S. M. Yen and H. J. Schmidt, "Monte-Carlo Solutions of the Boltzmann Equation for Heat Transfer Problems," 6th International Symposium on Rarefied Gas Dynamics, MIT, Cambridge, Mass., July 22-26, 1968.

3. Technical Reports

R-374 On the Equivalence of Regular Expressions, John P. Hayes (April, 1968).

R-375 A Study of Nordstrom-Robinson Optimum Code, Franco P. Preparata (April, 1968).

R-376 Essential Parameters in Sensitivity Analysis, Dennis F. Wilkie and William R. Perkins (April, 1968).

R-377 The Influence of Gas Composition and of Signal Waveform on Properties of the Plasma Display, Brij. Mohan Arora (May, 1968).

R-378 A Proposed Test of the Einstein Theory of Gravitation by Means of an Unshielded Orbiting Gyro Using Passive Telemetry, D. H. Cooper, G. R. Karr, J. L. Myers, D. Skaperdas (May, 1968).

R-379 Singular Perturbation Method in the Theory of Optimal Control, Peddupullaiah Sannuti (May, 1968).

R-380 Optimization of Systems Containing Discontinuous Elements, Robert Gerald Stefanek (May, 1968).

PUBLICATIONS

R-381 Optimal Test Sets for the Diagnosis of Multiple Output Combinational Nets, Richard D. Isenhart (May, 1968).

R-382 An Algorithm for Nand Decomposition of Combinational Switching Systems, Edward S. Davidson (May, 1968).

R-383 An Approximate Design of Optimal Regulations for High-Order Linear Plants, P. Sannuti and P. Kokotović (May, 1968).

R-384 Use of Switch Sensitivities to Provide Near Optimal Control, R. G. Stefanek and J. B. Cruz, Jr. (June, 1968).

R-386 Use of the Companion Transformation in Parameter Optimization and Adaptive Control, Dennis Frank Wilkie (July, 1968).

R-387 Tables of Hertzian Contact-Stress Coefficients, D. H. Cooper (August, 1968).

R-388 An Application of Ternary Logic to Logic Sequencing, Anthony S. Wojcik (July, 1968).

R-389 A Class of Optimum Nonlinear Double-Error-Correcting Codes, Franco P. Preparata (July, 1968).

(OVERLEAF BLANK)

1. SURFACE PHYSICS[†]

F. Propst
J. Burkstrand
T. Cooper
D. Edwards
C. Foster
M. Nishijima
G. Sakuri

1.1. Ion Emission from Surfaces Induced by Electron Bombardment

1.1.1. Introduction

The experimental apparatus for this study has been completed, and initial measurements on polycrystalline tungsten with oxygen adsorbed (O/W) have been made. The apparatus was discussed in the previous progress report and will not be described here. The polycrystalline-tungsten sample used in the present work was heated for about 15 hours to over 2200°K in oxygen at a pressure of approximately 1×10^{-6} Torr. Prior to each measurement, the sample was flashed two or three times to over 2200°K for approximately 3 sec. It is generally accepted that this procedure produces a clean tungsten surface. Our results indicate that this procedure may not be sufficient and that the treatment of the target subsequent to the initial cleaning can affect the cleanliness materially. Some of the principle features of the results obtained to date will be described below.

1.1.2. Identification of Ion Species

One of the unique features of this apparatus is that the mass to charge ratio as well as the energy distribution of the ejected ions can be determined. This can be understood from the following considerations.

[†]This work was supported by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

In a uniform magnetic field, B, an ion will perform a helical motion. The projection of this motion on a plane passing through the axis is given by

$$(M/R)(v\sin\theta)^2 = q(v\sin\theta)\cdot B, \tag{1}$$

where M is an ionic mass, q is the ionic charge, θ is the initial take-off angle of the ion, v is the velocity of the ion, and R is the radius of the helical motion. We know also, that

$$\tfrac{1}{2}Mv^2 = qV_T, \tag{2}$$

where qV_T is the total ionic energy, which includes the initial kinetic energy of the ion, the contact potential of the target, and the externally applied target bias.

Using Eqs. (1) and (2), the mass to charge ratio of the ion M/q is given by

$$M/q = \tfrac{1}{2}(R/\sin\theta)^2(B^2/V_T). \tag{3}$$

If the ion energy distribution for a particular ion is measured for two values of the target bias, the mass to charge ratio can be determined from the expression

$$M/q = \tfrac{1}{2}(R/\sin\theta)^2(B_1^2-B_2^2)/(V_1-V_2), \tag{4}$$

where B_1 and B_2 are the values of the magnetic field at the peak of the two distributions, V_1-V_2 the change in the applied bias, and R and θ are physical parameters of the apparatus.

Figure 1.1 shows the results of this type of measurement for O/W for the target bias set at V_1=3 volts, and V_2=5 volts. The mass to charge ratio calculated using Eq. (4) is 16.5. This confirms that the ions ejected from the O/W system by the incident electrons are O^+.

1.1.3. Time Dependence Measurements

The kinetics of the electron impact desorption, in the simplest case, is expressed by the equation

$$dN/dt = s(N)\gamma P - (j/e)QN, \tag{5}$$

where N is the surface coverage, s is the sticking coefficient, $\gamma = (MT)^{\frac{1}{2}}3.5\times10^{22}$, P is the pressure of the system, j is the electron current density, and Q is the total desorption cross section (ion and neutral). The first term represents adsorption, and the second term is the electronic desorption component.

Equation (5) can be solved easily, if s and Q are independent of N (small coverage), and the pressure is kept constant in time t. The solution is

$$N = N_1 e^{-t/\tau} + (e/jQ)s\gamma P(1-e^{t/\tau}), \tag{6}$$

where $N = N_1$ at $t = 0$, and

$$\tau = e/jQ. \tag{7}$$

If we can neglect the contribution due to adsorption, we have

$$N = N_1 e^{-t/\tau}. \tag{8}$$

If the ion current, i^+, is proportional to the surface coverage, N, we have,

$$i^+ = i_1^+ \cdot e^{-t/\tau}, \tag{9}$$

where i_1^+ is the ion current at t=0. The total desorption cross section Q can be determined from the time dependence of the ion current using Eq. (9), when adsorption is negligible.

If the initial coverage is zero, Eq. (6) becomes

$$N = \tau s \gamma P(1-e^{-t/\tau}), \tag{10}$$

and the ion current is given by

$$i^+ = i_e^+ q(1-e^{-t/\tau}). \tag{11}$$

Figure 1.2 shows an example of the ion current vs time for the O/W system at a constant pressure of oxygen. At t=0, the target was cleaned by flashing to over 2200°K. The curve consists of three regions, the initial portion of the curve with zero ion current, the intermediate region, and the saturation region which corresponds to the equilibrium condition dN/dt=0. The time interval of the initial portion of the curve (delay time) is proportional to the inverse of the pressure. This type of curve indicates that there are two adsorption states (state 1 and state 2) on the surface. State 1 corresponds to a state with very-low ion-desorption cross section and high binding energy, and State 2 to a state with the higher ion-desorption cross section and lower binding energy. State 2 begins to fill with oxygen only after State 1 is almost full. State 1 is

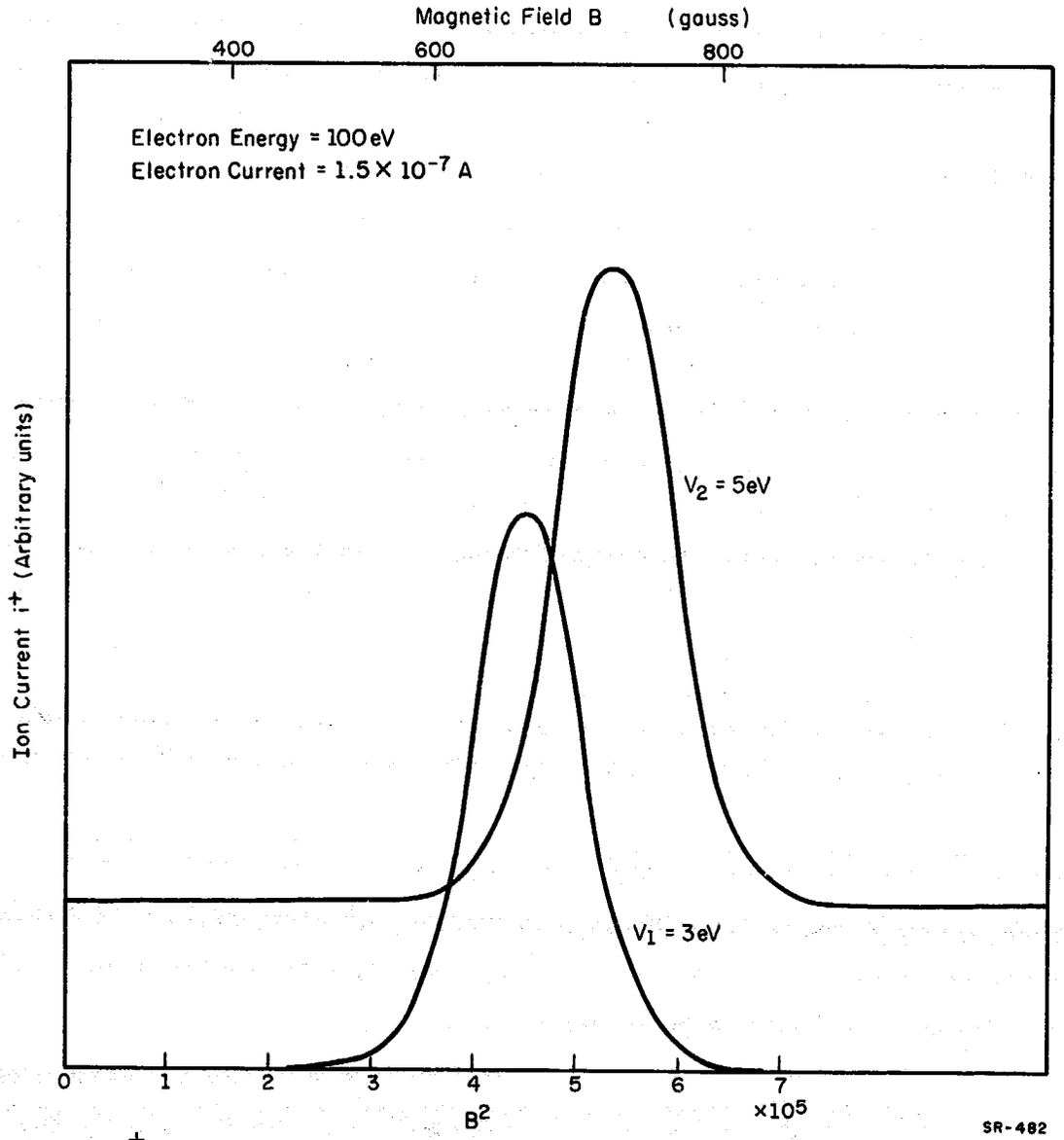


Fig. 1.1. O^+ ion-energy distributions for two values of the target bias. The charge-to-mass ratio calculated from the shift in the distribution is 16.5.

(OVERLEAF BLANK)

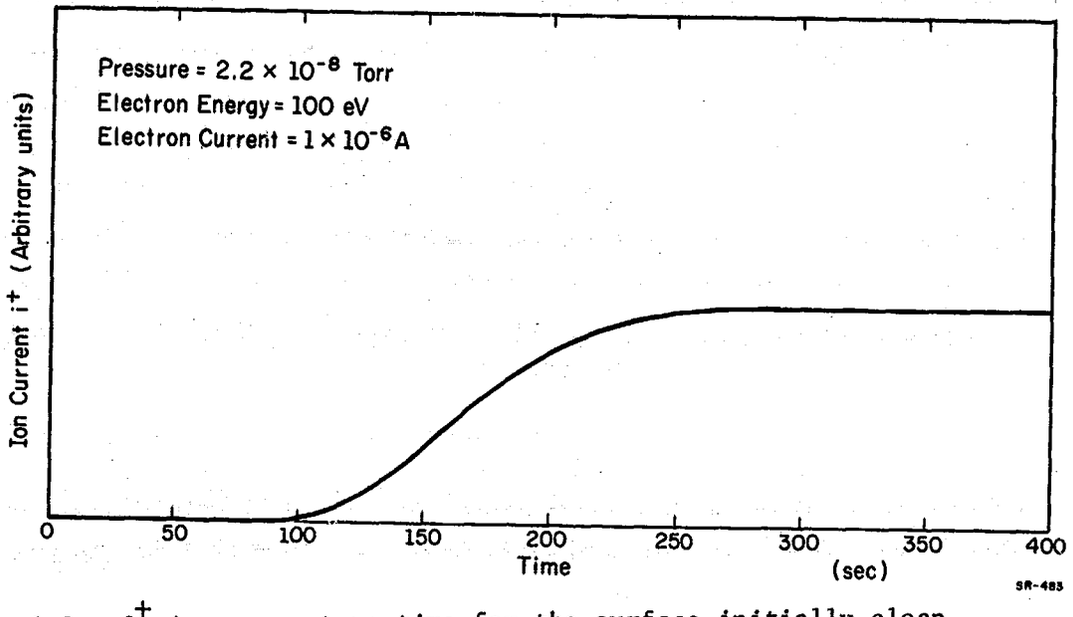


Fig. 1.2. 0^+ ion current vs time for the surface initially clean.

(OVERLEAF BLANK)

found to have an ion desorption cross section of at least a factor of 10^3 smaller than that of State 2.

Figure 1.3 shows a typical plot of ion current vs time for a surface with full coverage initially and at a low oxygen pressure. The curve shows an initial sharp decay, an intermediate region, and a saturation region. A logarithmic plot of this curve does not yield a straight line. This can be attributed to either a variation of the total desorption cross section, Q, or of the sticking coefficient, s, with the surface coverage, N. Thus, the simple analysis reflected in Eqs. (6)-(11) does not hold. A computer fit of the experimental data to Eq. (5), with the functional dependence of s and Q on N included, will be attempted.

Figure 1.4 shows logarithmic plots of desorption curves for different electron energies. The curves are normalized at t=0. The initial slopes of the curves in Fig. 1.4, which are proportional to the total desorption cross section at full coverage, shows that the cross section increases substantially from Ve=30 eV to about Ve=100 eV, and that it saturates between Ve=100 eV and Ve=500 eV.

At the equilibrium dN/dt=0; therefore, Eq. (5) gives

$$s(N) = jQN/e\gamma P \propto Q_i^+/Q^+, \qquad (12)$$

where Q^+ is the ion-desorption cross section. The variation of the sticking probability s with the surface coverage, N, can be measured if the total desorption cross section Q and the ion-desorption cross section, Q^+, are independent of the surface coverage N. Figure 1.5 shows the

experimental results of this type of analysis. The general shape of the curve agrees with the results reported by other experiments.

1.1.4. Ion-Desorption Cross Section

The ion current is related to the electron current i^- by

$$i^+ = i^- N Q^+, \tag{13}$$

where Q^+ is the ion-desorption cross section. Figure 1.6 shows the ion-desorption cross section as a function of electron energy for O^+ ions from the O/W system. The measurement was made with the surface fully covered. This energy dependence is similar to that of the total desorption cross section.

In the above experiments, the linearity of the ion current with the incident electron current was confirmed, showing that the ion emission is a single-step process involving only one electron.

1.1.5. Ion-Energy Distribution

The energy distribution of O^+ ions from the state 2 is given in Fig. 1.7. The distribution has a maximum at about 7.5 eV, and a half width of 2.8 eV. A small electron current of 2×10^{-7} A was used to minimize the electron-impact desorption during a single sweep of the ion-energy distribution. These data will be used to obtain information on the interaction potential of ions with the surface.

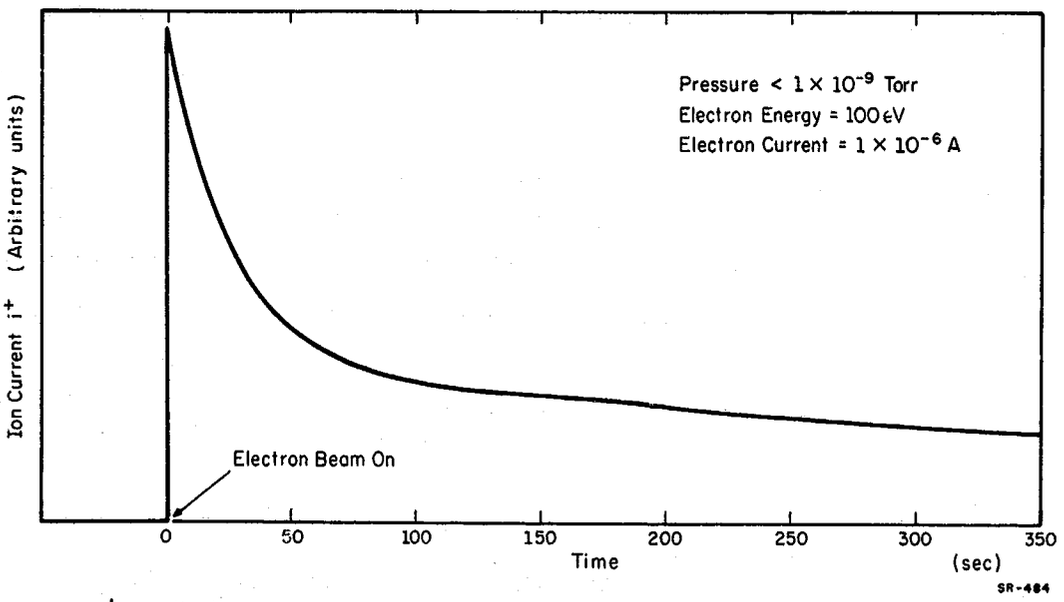


Fig. 1.3. 0^+ ion current vs time for the surface initially covered to approximately full coverage and at low oxygen pressure.

(OVERLEAF BLANK)

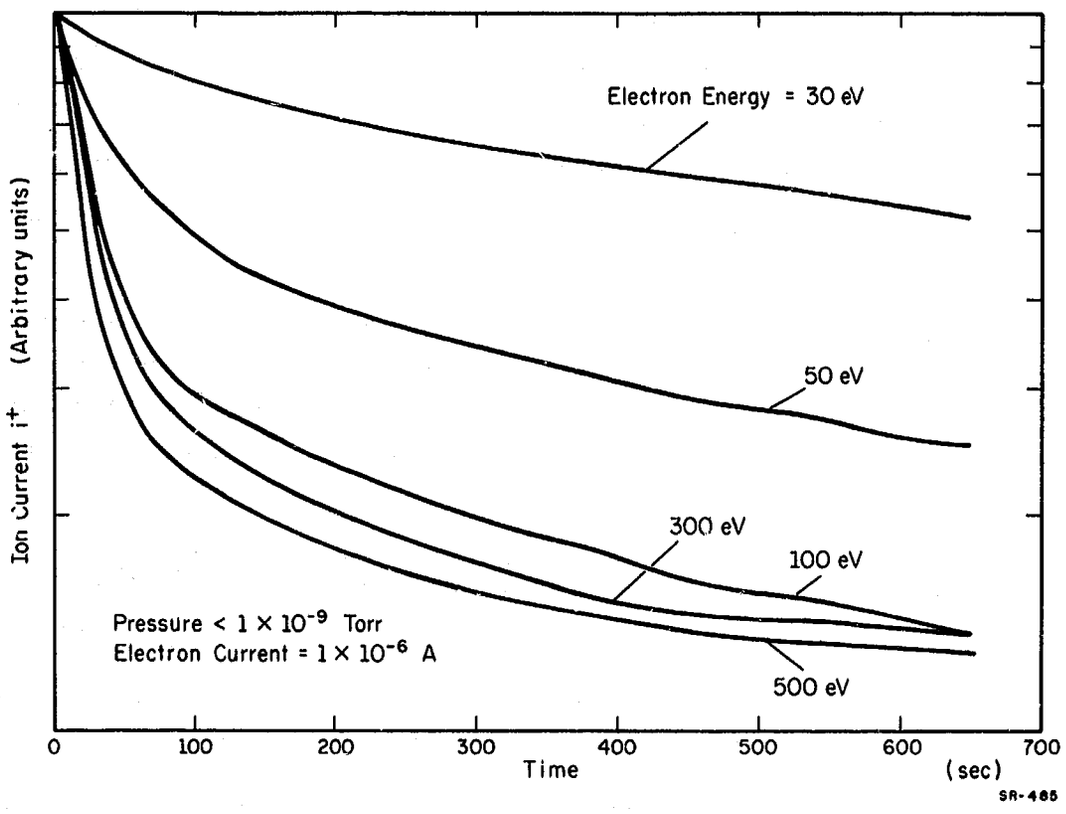


Fig. 1.4. Logarithmic plots of ion desorption curves (Fig. 1.2) for various electron energies. The curves are normalized at t = 0.

(OVERLEAF BLANK)

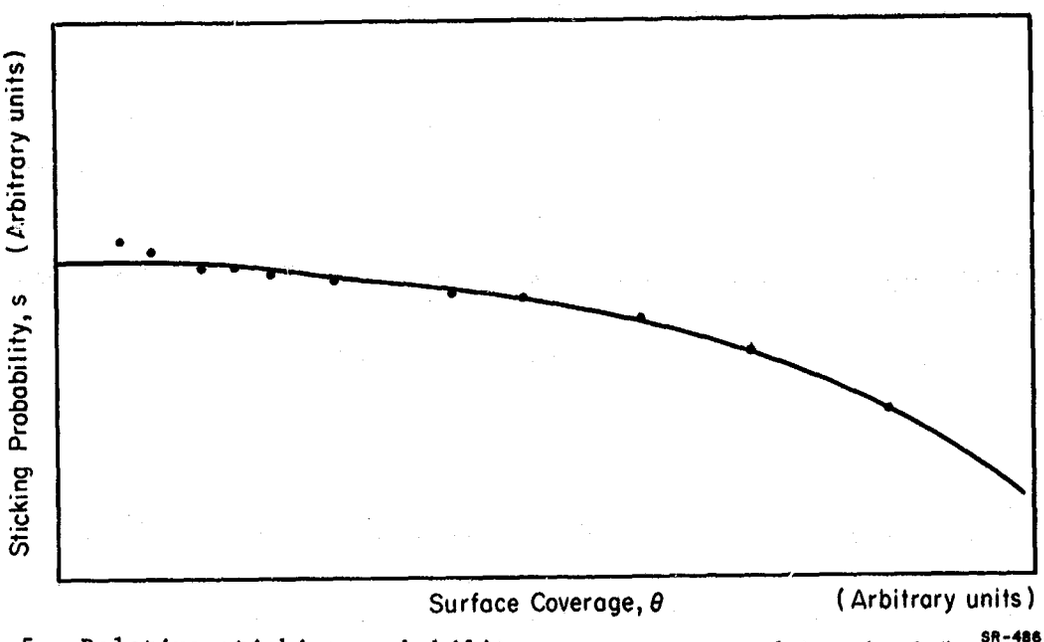


Fig. 1.5. Relative sticking probability vs coverage as determined from the equilibrium ion current at various bombardment electron currents. SR-486

(OVERLEAF BLANK)

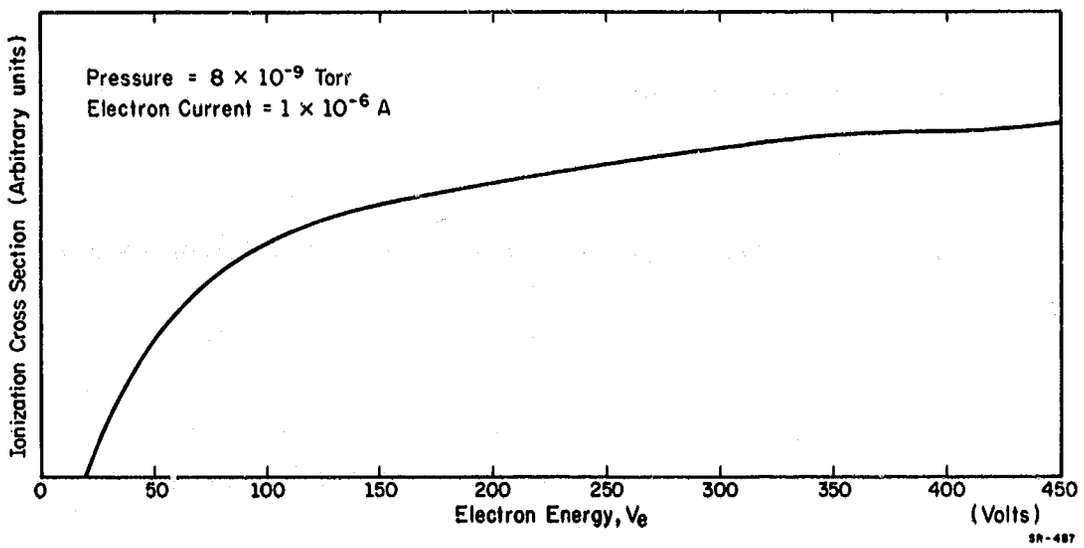


Fig. 1.6. Relative ion-desorption cross section as a function of electron energy.

(OVERLEAF BLANK)

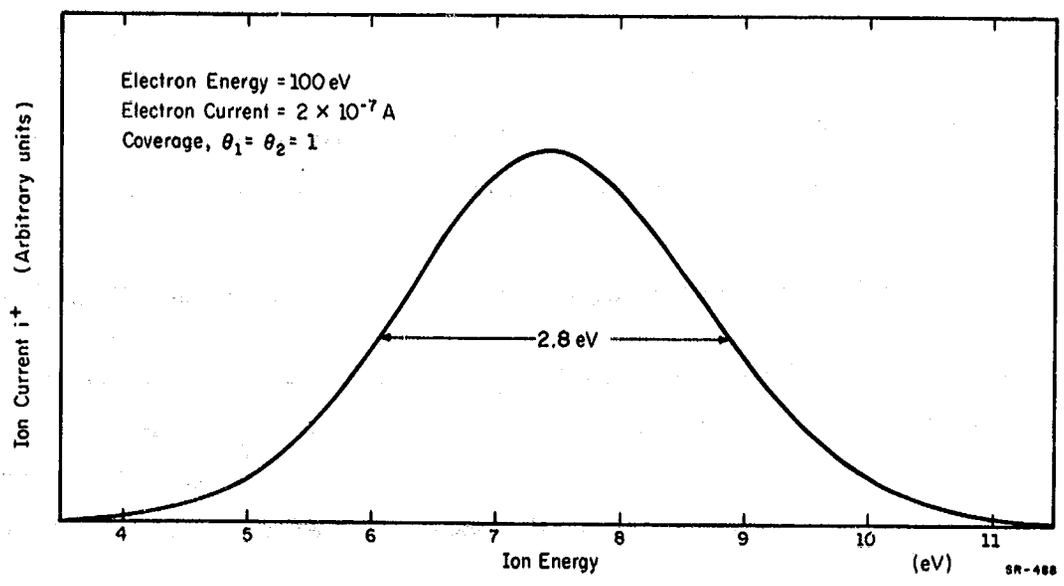


Fig. 1.7. O^+ ion energy distribution from fully-covered tungsten surface.

(OVERLEAF BLANK)

1.2. Study of Polycrystalline Platinum Using Low-Energy Electrons

1.2.1. Introduction

Since platinum has particular importance as a catalytic agent, the low-energy electron spectrometer described in previous progress reports was applied to the study of polycrystalline platinum. Polycrystalline material was chosen for this study since this is the type of material most commonly used in practical applications and since it was felt that many of the important features could be determined using the polycrystalline material. It has been reported in the literature that platinum is extremely difficult to use in surface work since clean surfaces are so hard to produce and maintain. Since the electron spectrometer is particularly sensitive to contaminants, it would be interesting to apply this technique to a system known to have contamination problems. Some of the results and tentative conclusions drawn from these measurements will be described below.

1.2.2. Experimental Results

Two types of measurements have been made on this system. The first is the measurement of the elastic-reflection coefficient, and the second is the measurement of the energy distribution of secondary electrons. In this work, the elastic-reflection coefficient is defined as the ratio of the number of electrons leaving the sample having suffered no energy loss and traveling in the specular direction to the number of electrons striking the surface. This coefficient is measured as a function of the energy of the electrons striking the surface. The energy

distribution of secondary electrons can have a structure which is related to the band structure of the solid and to contaminants on the surface of the sample. In particular, the vibrational states of gases adsorbed on surfaces can be detected by this technique.

The platinum sample was cut from a sheet of high-purity rolled foil. It was cleaned chemically and installed into the system. The sample was then heated by electron bombardment. The energy distribution of secondary electrons showed a satellite peak at a loss energy of approximately 75 meV for all heating temperatures up to approximately the melting point. The sample was heated in the presence of oxygen and then in vacuum in an attempt to remove the contaminant producing this peak. We were unable to eliminate the peak by any combination of heating and oxidation. An ion-bombardment apparatus was installed into the system in an effort to clean the sample by ion bombardment. However, it was again found that it was impossible to remove the satellite peak by any combination of bombardment, heating, and oxidation. This is taken as strong evidence that the procedures employed were not capable of producing a clean surface. In previous experiments with tungsten, such a result was interpreted as indicating either a carbon- or silicon-bearing contaminant. During these experiments, the pumping station for the apparatus was an oil-diffusion pump using silicon pumping fluid. We feel it very likely that the sample was contaminated with silicon from the pumping fluid. Such a contaminant would be expected to materially affect the activity of the surface towards the adsorption of gases. Indeed this

was found to be the case. We found that the surface showed very little activity towards the adsorption of hydrogen, and nitrogen after any of the processing described above.

Carbon monoxide was adsorbed only after the argon-ion bombardment and oxidation of the surface. The adsorption of carbon monoxide produces a molecular vibration at 258 meV. This agrees well with the value reported by Eischens for CO adsorbed on powdered platinum. It is interesting to note that in the present experiments, the adsorption almost definitely occurs on a contaminated sample.

In the process of attempting to observe vibrational levels of adsorbed gases, several interesting results were noted in the elastic reflection coefficient and in the energy distribution of scattered electrons. These results will be described in the following paragraphs.

Figure 1.8 shows the energy distributions of secondary electrons from the platinum sample for several primary energies and as a function of the power used in annealing the sample. It will be noted that the unannealed sample shows no structure except the elastic peak and the broad low-energy peak of secondary electrons. As the sample is heated to higher and higher temperatures, peaks begin to appear at approximately 1, 3, and 7 eV secondary energies, the peaks becoming more and more pronounced as the annealing temperature is increased. In addition, a peak begins to appear at an energy of approximately 7 eV below the elastic peak as the target is heated. This peak is at a constant energy below the primary peak regardless of the value of the primary energy. This

indicates that there is a seven-volt excitation in the sample which is excited by the primary electrons.

Figure 1.9 shows the elastic reflection coefficient as a function of the annealing power. A similar result is noted. The curve for the unannealed target shows no structure except for the sharp peak at low energies, 3, 8, 17, and 20 eV. The appearance of this structure was interpreted as being related to the recrystallization of the sample upon annealing. To test whether this recrystallization was due to bulk annealing or to surface annealing, a set of experiments similar to those described were conducted after argon-ion bombardment. The sample was bombarded with argon ions of 300 eV kinetic energy for 15 minutes, a current density of approximately 30 $\mu A/cm^2$. Figure 1.10 shows the elastic reflection coefficient after this sputtering. It is apparent that it is relatively free of structure and is similar to the curve for the unannealed target. Figure 1.11 shows the energy distribution after the sputtering. Again, the curve shows essentially no structure and is very similar to the unannealed sample. Figure 1.12 shows the energy distributions of secondary electrons for several primary energies after annealing the sputtered sample to approximately 40 W of heating power. It is evident that the sample shows structure that is similar to that observed prior to bombardment but after annealing. However, the peak at approximately 3 eV secondary energy is substantially stronger than prior to the sputtering. In addition, this set of curves shows that the peak at 7 eV below the primary is at a constant energy loss and is strongest when it occurs at a secondary energy of approximately 7 eV.

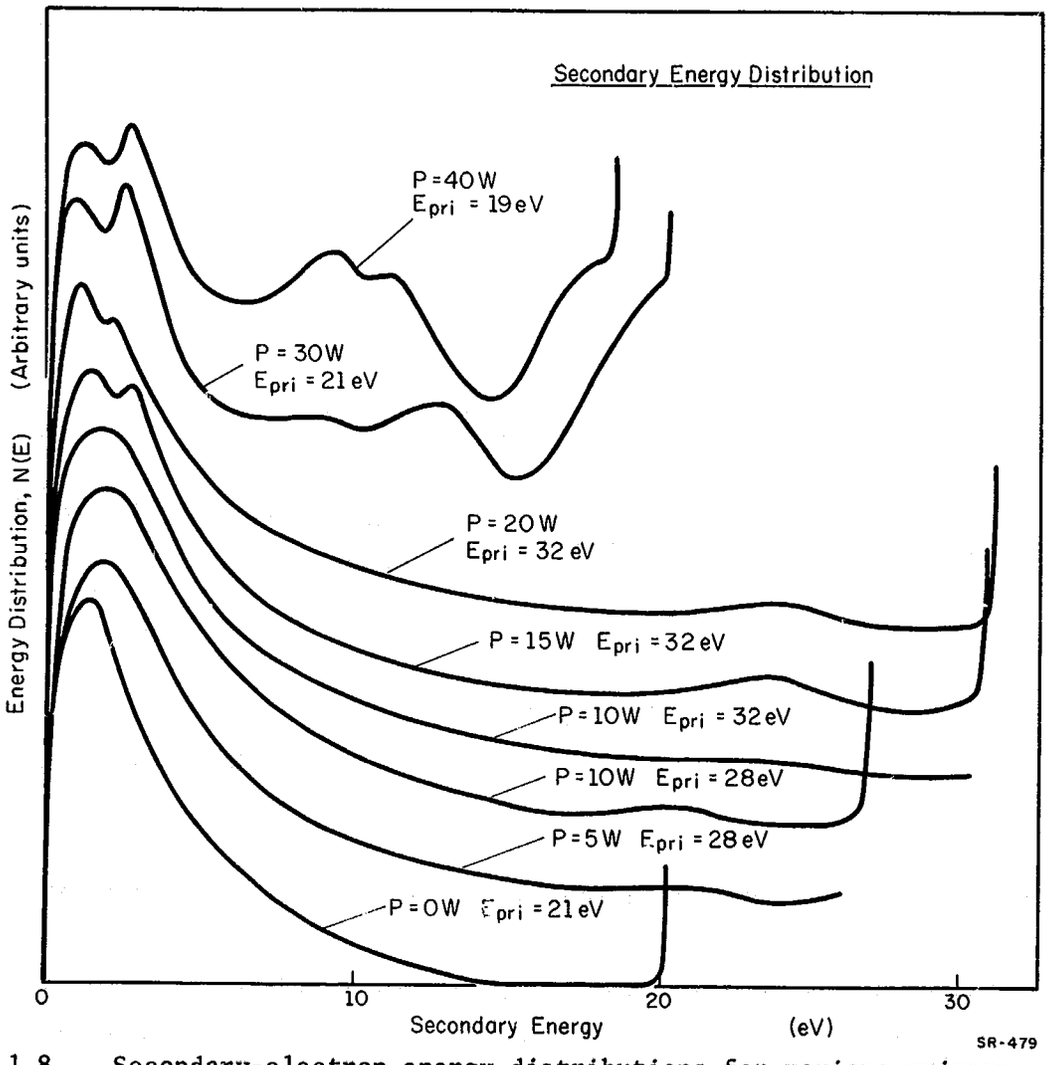


Fig. 1.8. Secondary-electron energy distributions for various primary energies and annealing powers.

(OVERLEAF BLANK)

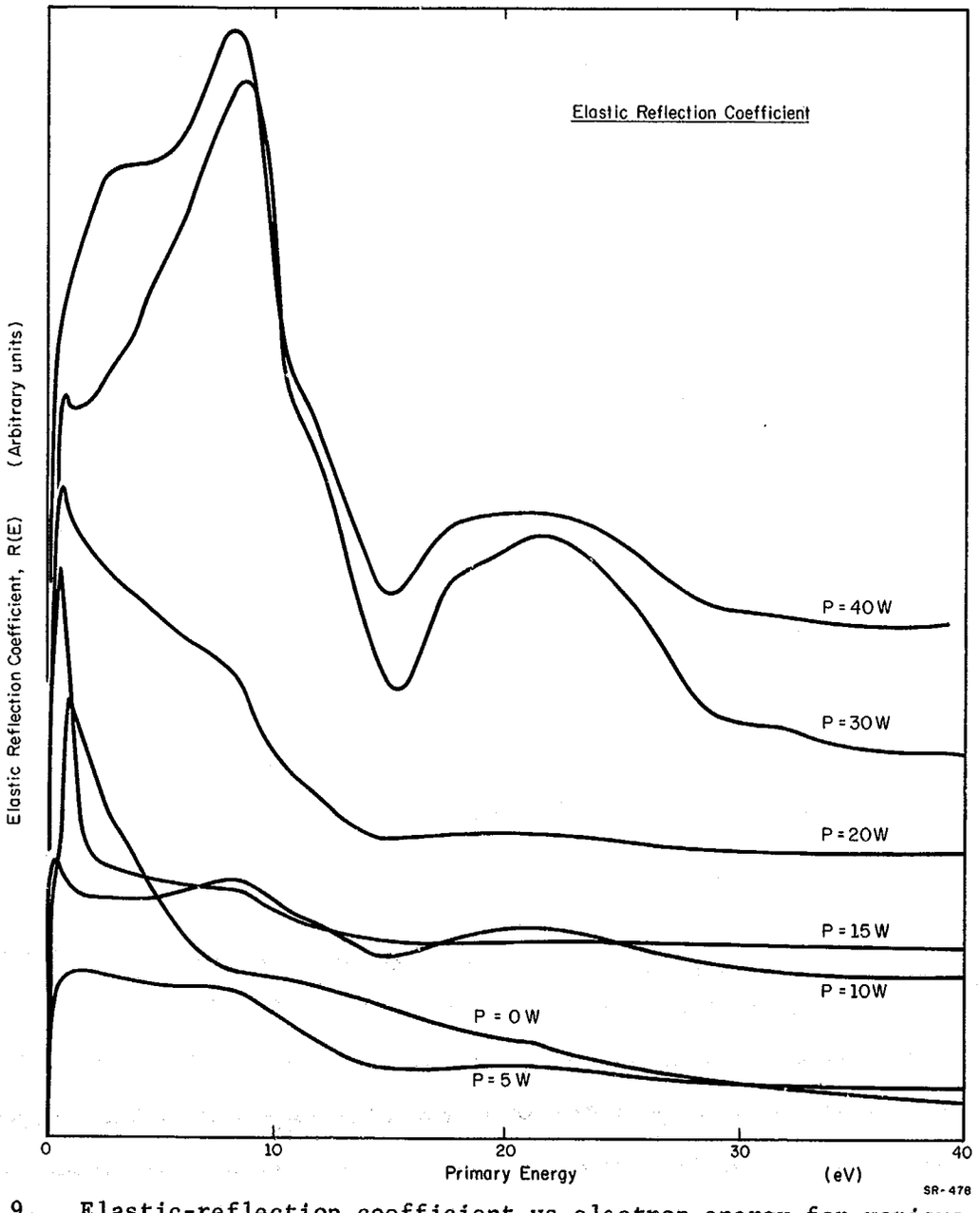


Fig. 1.9. Elastic-reflection coefficient vs electron energy for various annealling powers.

(OVERLEAF BLANK)

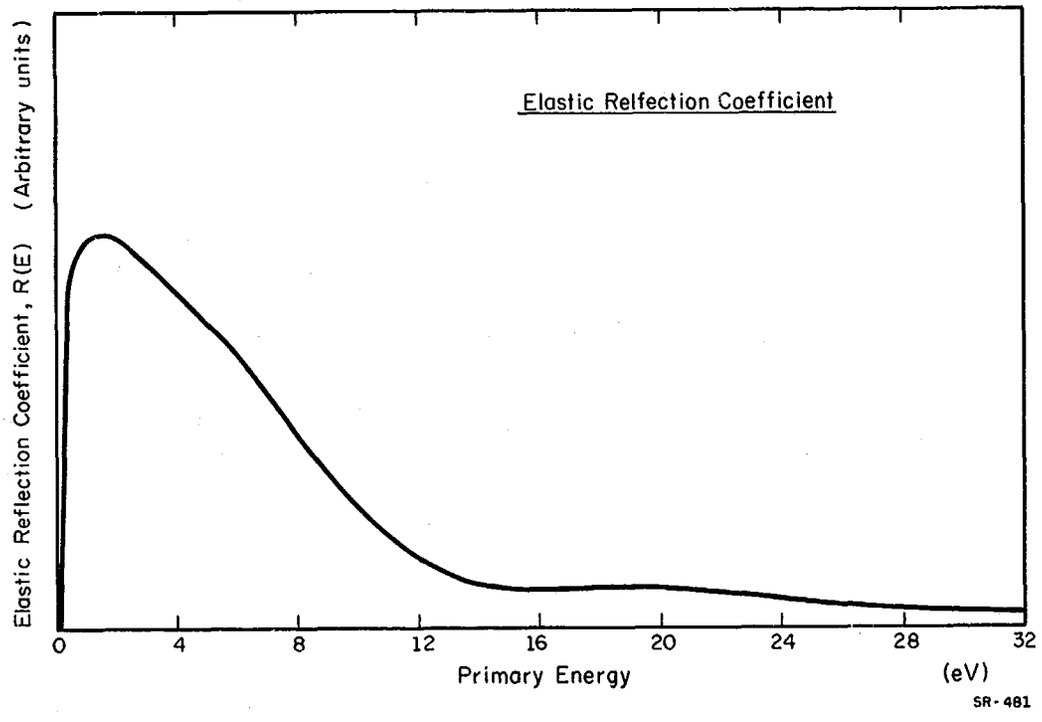


Fig. 1.10. Elastic-reflection coefficient vs electron energy after argon-ion bombardment and prior to annealing.

(OVERLEAF BLANK)

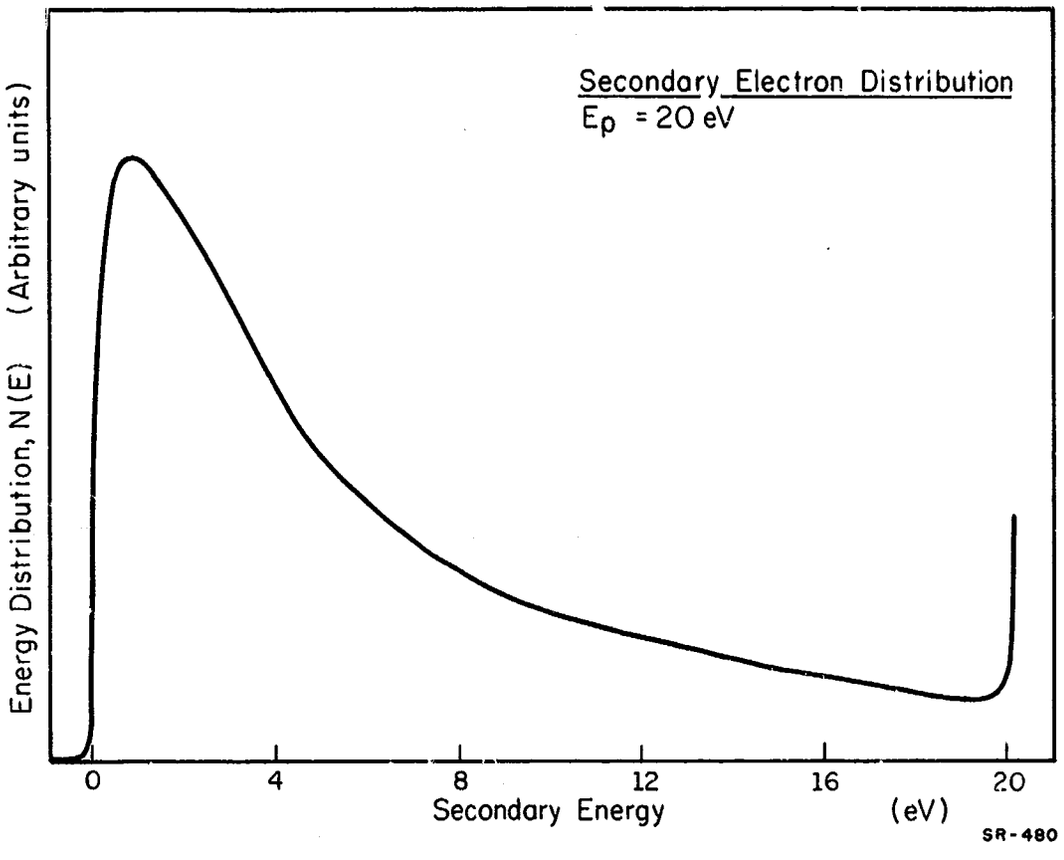


Fig. 1.11. Secondary-electron energy distribution after argon-ion bombardment and prior to annealing.

(OVERLEAF BLANK)

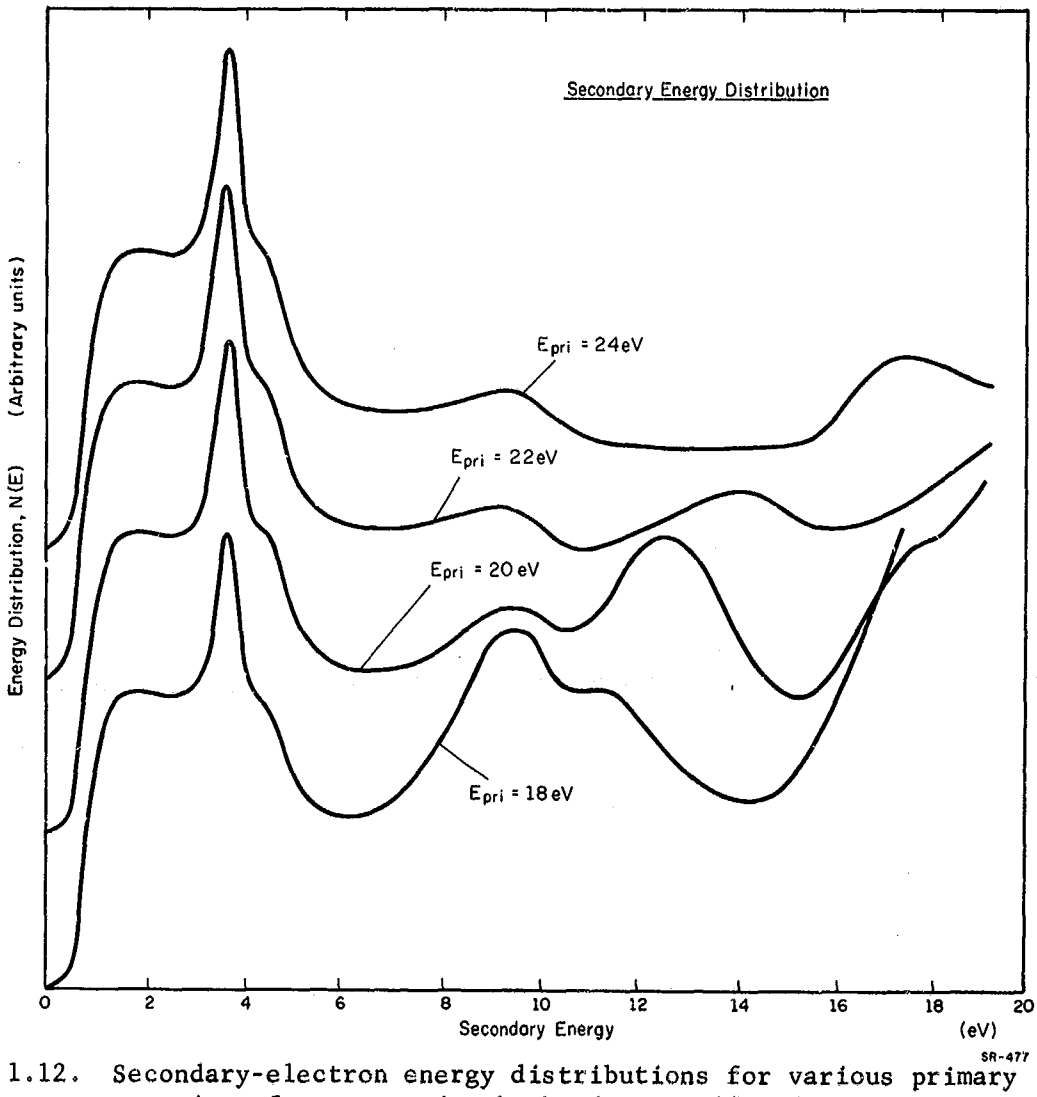


Fig. 1.12. Secondary-electron energy distributions for various primary energies after argon-ion bombardment followed by annealing at 40-W annealing power.

(OVERLEAF BLANK)

These results show that the structure observed in the two types of experiments are due to recrystallization of the sample, and that the region of importance is very near to the physical surface, since the ion bombardment damages only the first few atomic layers. As to the source of the observed structure, the peak at 7 eV loss energy is most likely a surface plasmon, and the peaks at 1 and 3 eV are likely related to the resonance transmission of secondary electrons through the surface barrier. These types of structure have been observed with tungsten and have been interpreted previously in this manner; however, the validity of these interpretations is not at all well established, and we must wait further experimental and theoretical work before it is established firmly.

1.2.3. Summary

We have attempted to observe the vibrational states of gases adsorbed on polycrystalline and recrystallized platinum. Of the gases tried (hydrogen, nitrogen, carbon monoxide, and oxygen), only carbon monoxide produced well-defined excitations. The energy of vibration agrees well with that observed previously and interpreted in terms of the bridge-bonded CO. However, the sample indicated that there was some contaminant which could not be removed, and this probably materially affected the adsorption characteristics. The recrystallization could be monitored by either the elastic reflection coefficient or the energy distributions of secondary electrons, and this technique is sensitive to disorder in the first few atomic layers. However, the detailed interpretation at this time, of the structure in these two types of experiments

are again deferred because of the presence of the contamination and lack of theoretical understanding of the electron-scattering process.

1.3. Energy-Angular Distributions of Secondary Electrons

Several difficulties with the vacuum system for the experiment developed during the early part of the period for this report. First, a leak developed in the pumping station. It was necessary to dismantle the system in order to repair this leak. Upon doing this, it was found that the two cold traps were heavily contaminated with mercury, and the copper support block in the refrigerated baffle showed signs of serious almagamation. The traps were cleaned by heating in air to 350°C for 24 hours with subsequent firing in vacuum to 750°C for two hours. The baffle and pumps were cleaned with organic solvents and vapor blasted. The copper block in the baffle was nickel plated, and the baffle was welded directly to the diffusion pump. The flat platinum gasket flanges on the baffle and on the trap directly above the baffle were replaced with Conflat* flanges, and the copper gaskets used in these flanges were nickel plated. The previous flange design which was used had caused problems on numerous occasions and were probably the source of the leak mentioned above.

The vacuum performance of the system before the changes described above was relatively good; however, mass-spectrometric analysis indicated that the major portion of the background gas was hydrogen. The pumping

*Varian trade name.

speed of the mercury pumping system was not sufficiently high to adequately handle this hydrogen load. A titanium-sublimation pump was constructed and added to the system to supplement the mercury pumping system.

After making these changes and adding the sublimation pump, the system was baked at 250°C for ten hours, followed by a bake of the traps with the main valve closed at 250°C for ten hours. After this processing, the system pressure was 1×10^{-10} Torr.

Some preliminary measurements of elastic and inelastic electron scattering from 100 tungsten have been made. Figure 1.13 shows the elastic reflection in the 110 plane for primary electrons of 112 eV energy. the major peak in this curve is the 11 defraction beam. The two satellite peaks are probably due to contamination but could not be removed by flashing the target to approximately 2500°K. A full diffraction pattern was taken by observing the diffraction beams as the target was rotated about the axis of the apparatus. This resulted in a distorted square pattern with large intensity variations in equivalent peaks (e.g., the 10 beam might be of the order of ten times as intense as the 01 beam). The distortion of the pattern could be explained in terms of the lack of complete alignment of the electron gun, collector, and target; however, the strong intensity differences in equivalent beams and satellite peaks such as those shown in Fig. 1.14 indicate a damaged or highly contaminated surface.

The energy distribution of secondary electrons leaving the surface in the direction of the 11 diffraction beam shown in Fig. 1.12 is given in Fig. 1.13. There are loss peaks at 9 and 24 eV loss energy.

These peaks likely are due to surface and volume plasmons respectively. Figure 1.15 shows the angular distribution of the 24 eV loss. It will be noted that structure does exist in the angular distribution of the electrons which have excited the 24 eV mechanism; however, there does not appear to be a simple relationship between this structure and the structure in the angular distribution of elastically scattered electrons. In particular, there seems to be no correlation with the angular position of the 11 beam. Both the energy of the surface plasmon and the angular distribution of electron which have excited the surface and bulk mechanism are very likely affected by the contamination on the surface.

Because of the various indications of contamination or damage discussed above, we decided to cut another crystal paying more attention to the mechanical finishing and cleaning in an effort to produce a cleaner and more perfect surface. A 0.025-in slab was EDM machined from a tungsten ingot. This slab was then polished with diamond grit. Periodically during this polishing, the surface was electrochemically etched in a 1% NaOH solution. The sample was then electropolished in the NaOH solution until all surface damage due to the mechanical polishing should have been removed. The target was then fired in vacuum at 2200^{o}K for 24 hours and then fired in approximately 1×10^{-6} Torr of oxygen at 2200^{o}K for 24 hours. This crystal has been installed in the apparatus and the electric gun, collector, and target aligned as accurately as possible (approximately within 2^{o}). Initial measurements have been improved; however, there still appears to be a problem with surface contaminants. These initial results

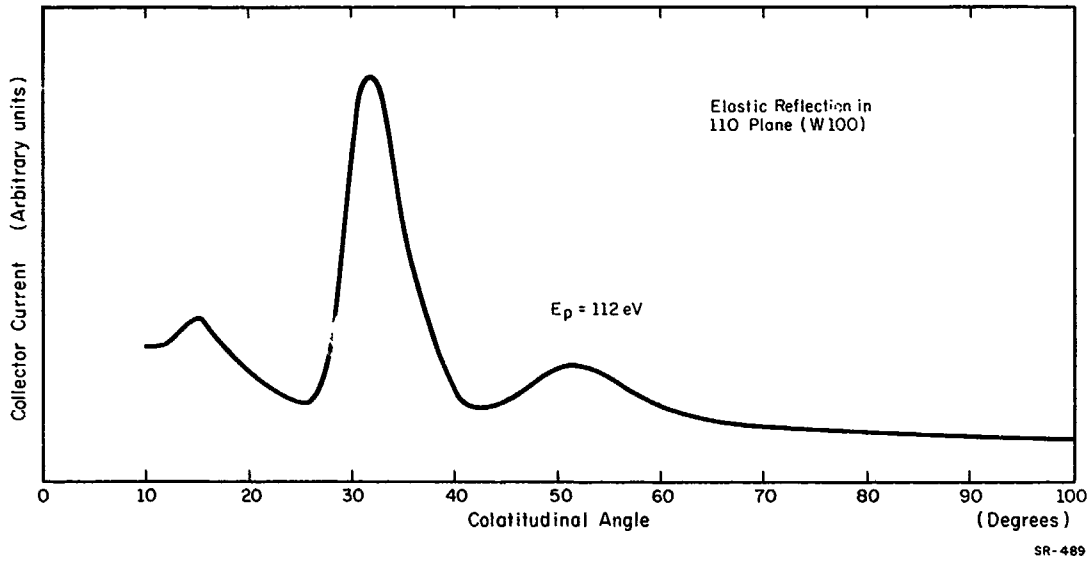


Fig. 1.13. Diffraction peak along 11 direction produced by a 112 eV normally-incident beam, 100 tungsten surface.

(OVERLEAF BLANK)

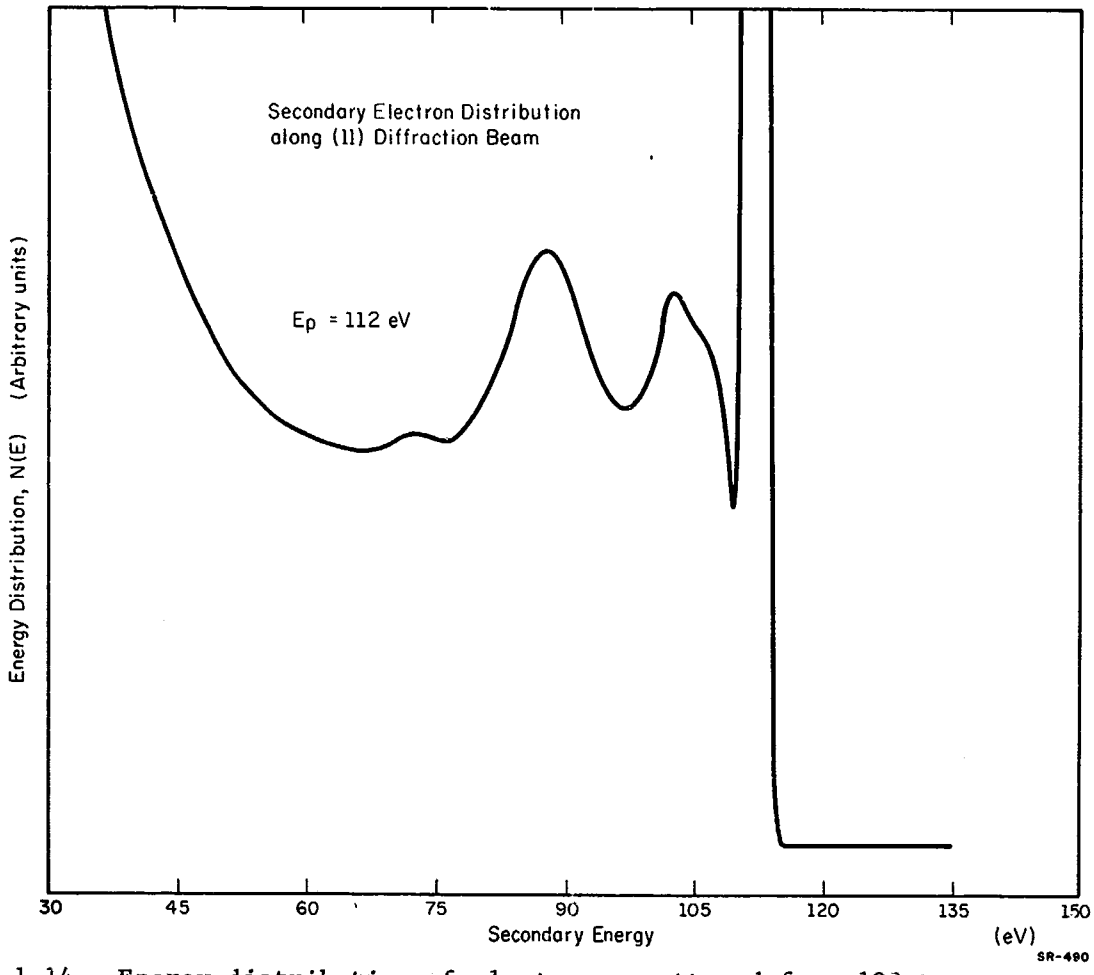


Fig. 1.14. Energy distribution of electrons scattered from 100 tungsten by 112 eV electron beam.

(OVERLEAF BLANK)

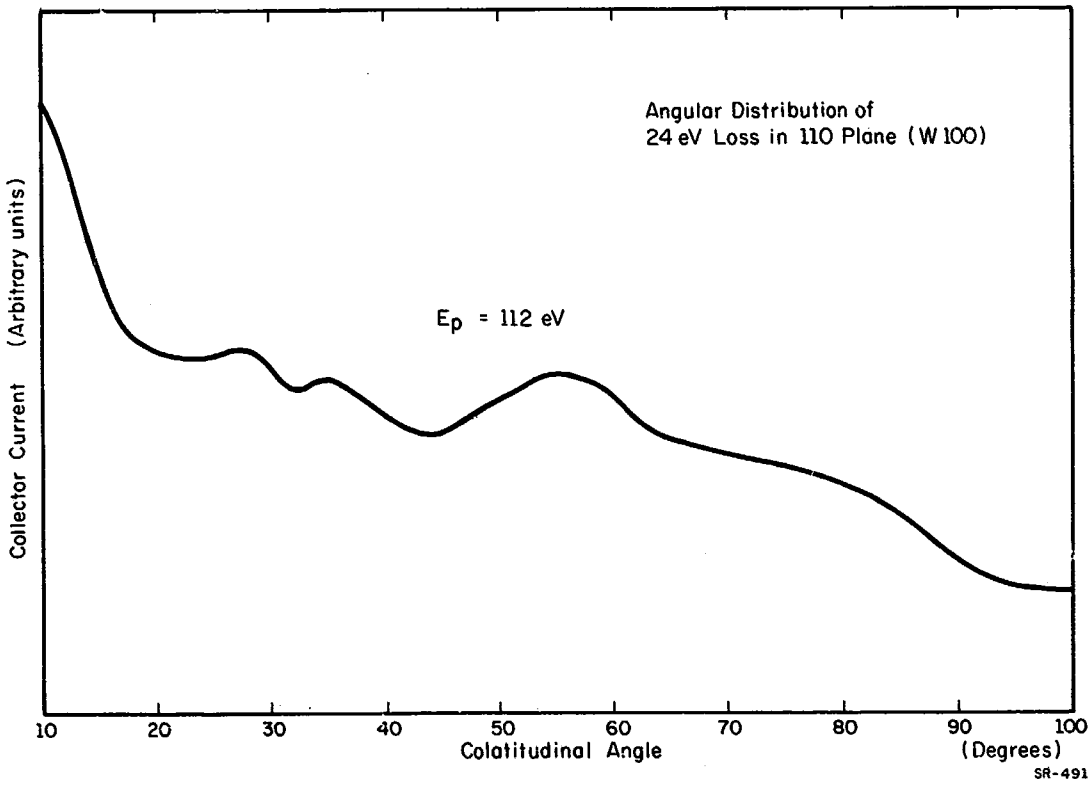


Fig. 1.15. Angular distribution along a 11 direction of a 100 tungsten surface of a 24 eV loss peak produced by a 112 eV normally-incident electron beam.

(OVERLEAF BLANK)

are an improvement over those obtained from the previous sample, and it is hoped that additional vacuum processing will be capable of generating a clean surface.

(OVERLEAF BLANK)

2. INFRARED CONVERTER† AND APPLIED PHYSICS

G. E. Anner
R. C. Birtcher
W. P. Bleha
D. E. Coad
W. D. Compton
J. W. Culton
S. W. Depp
J. T. Jacobs
R. L. Jiminez
D. L. Miller
R. N. Peacock
O. G. Reilly
J. E. Robinson
B. Tao
F. Steinrisser
A. C. Tulumello

2.1. Introduction

The basic principles of the direct image infrared converter discussed in this section were presented in earlier reports. Simply, the proposed device would be a planar thin-film arrangement utilizing the tunneling of optically-created minority carriers (holes) from a narrow-band-gap semiconductor into a broad-band-gap semiconductor where recombination with majority carriers (electrons) would generate visible light. The phosphor film which has been explored in detail is CdS, and the small-band-gap semiconductor used in most of the experimental structures has been germanium. For possible use at longer wavelengths, films of PbTe and PbS have also been examined. To better understand the operation of this device, work has also been done on single elements fabricated on single-crystal material. Since tunneling through insulating barriers is basic, some calculations and experiments on this subject are underway.

2.2. CdS Films by Vacuum Deposition

The latest CdS evaporations from the new enclosed chamber,

†This work was supported by Army Contract DAAK-02-67-C-0546 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

first used in the previous reporting period, have produced films showing the characteristic form of the edge photoluminescence (77°K) of pure CdS single crystals. The zero- and one-phonon peaks are very broad compared to a good CdS single crystal, but are close to the same wavelengths. The third peak is a shoulder on the curve. A photoluminescence spectrum is presented in Fig. 2.1. The film was evaporated on a 200°C substrate and given no further treatment. The room-temperature resistivity of 10^4 Ω-cm is similar to that of previous films.

The luminescent intensity is good and is comparable to films that have had a post-deposition heat treatment in H_2S. Although there is a faint luminescence peak in the blue near the free-exciton peak at 0.4875 μm, it is too broad to determine whether the luminescence is the high- or low-energy exciton series of CdS. The particular films studied were evaporated on fused-quartz substrates and the luminescent measurements were taken from the top of the three-micron-thick film. The photo luminescence from the face on the fused quartz was not as bright and was yellow green in color. In some samples no photoluminescence was observed from this face. Apparently there is a transition region of CdS between the amorphous substrate and more perfectly crystallized upper layers of CdS.

Since the active area of the device for generating luminescence will almost certainly be the junction between the CdS film and a Ge or Ge-insulator substrate, it is essential that measurements be made, first,

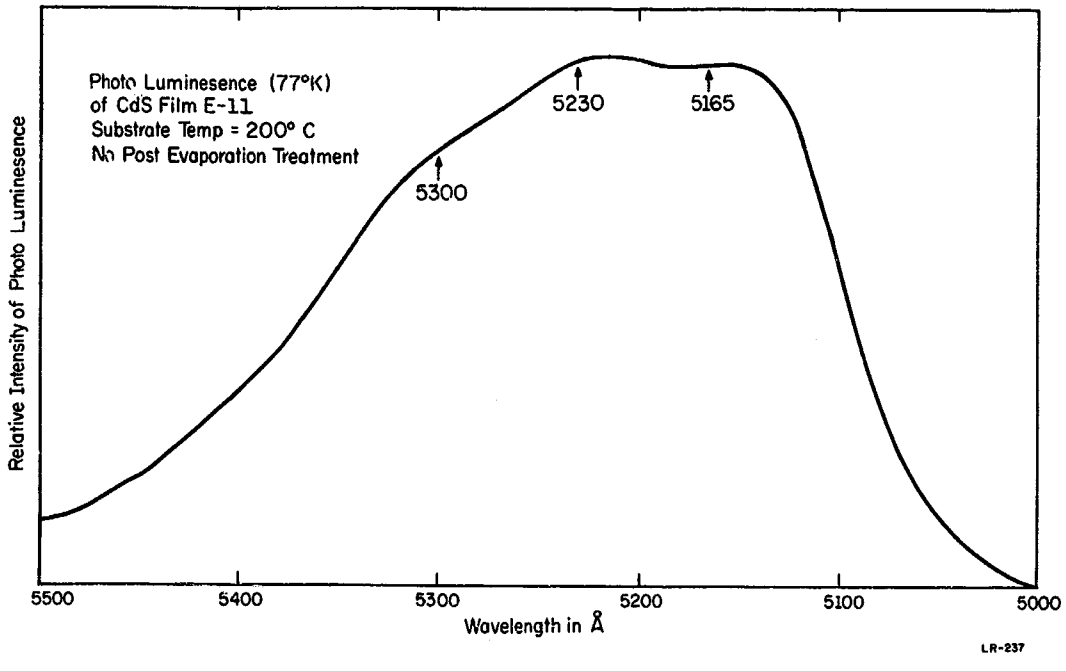


Fig. 2.1. Photoluminescence spectrum for a CdS film, as deposited.

(OVERLEAF BLANK)

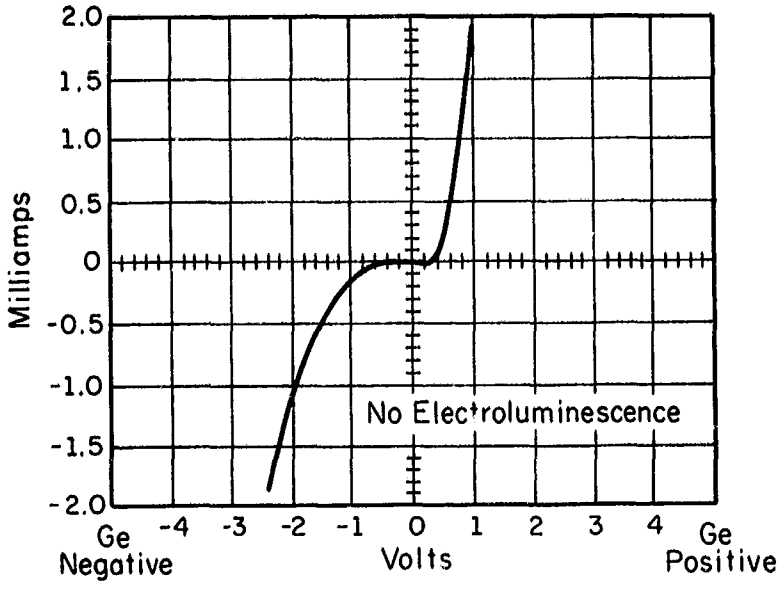
in finding the width of such a transition region and, secondly, minimizing it. Since crystalline Ge will be used as the substrate in the device, epitaxy may occur, although the crystal structures are dissimilar. It has been noted that the CdS evaporation rate decreases as a function of evaporation time. Possibly using a slower initial evaporation rate will lead to better initial crystallization. This will be investigated further during the next period.

The evaporation of CdS crystals doped with $CdCl_2$ was attempted, using the same procedure as that for pure CdS. The films had an appearance similar to the pure films but showed no photoluminescence at either $300^{o}K$ or $77^{o}K$. The resistivity of the films has not yet been measured, but it appears that post treatment will be necessary to produce luminescence.

2.3. CdS-Ge Heterojunctions

CdS films have been evaporated on air-cleaved, degenerate, n-type germanium and on similar germanium with a thin film of SiO evaporated on the cleaved face. CdS films deteriorate in the atmosphere and peel off the Ge surface. This could largely be prevented by baking the samples in an H_2S-HCl-H_2 atmosphere at $450^{o}C$. After the baking process, the films showed bright green photoluminescence at $77^{o}K$. It has not been possible to get a measurement of the resistivity perpendicular to the substrate because of the difficulty in obtaining an ohmic contract capable

of withstanding the 450°C bake. The resistivity of the CdS film parallel to the substrate is in the order of 10^6 Ω cm after the baking process. It is also necessary to obtain a transparent ohmic contact on the top of the CdS film in order to see the electroluminescence (EL) from the CdS-Ge junction. This is obtained by the glow discharge in air of the CdS film surface. This allows an ohmic contact to be made by many metals which normally make a rectifying contact. [1] After the discharge, a thin gold film is evaporated on the film. No attempt has been made to optimize the discharge parameters or find the effect on the other CdS film properties. The samples are then characterized by the current-voltage characteristic and observation of any EL. Typical I-V characteristics for samples that showed EL are shown in Fig. 2.2. There are two distinct characteristics, one corresponding to a state in which EL is seen, and the other showing no EL. The unit can be changed from one state to another by increasing the dc bias voltage, which apparently causes a forming process or discontinuity at some point in the junction which changes the characteristic. Thus far, no units have shown EL of enough intensity to record the spectrum, but visual observations have shown distinct points of both green and yellow luminescence at 77°K. No value can be given for the actual voltage across the junction region, and the voltage shown in Fig. 2.2 is that applied across the entire unit. Reproducibility has been low with most samples showing no EL. Also, samples breaking down to either open or short



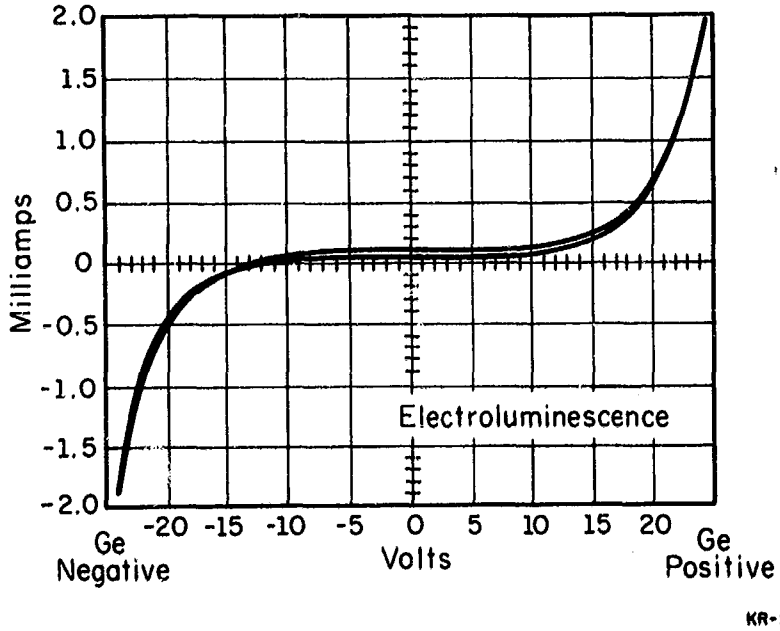


Fig. 2.2. Current-voltage characteristics for CdS-Ge junctions.

(OVERLEAF BLANK)

circuits at low voltages are common. The effect of the baking process on the Ge-CdS interface is critical in determining the properties of the system, and thus this must be controlled very carefully.

W. Bleha
D. Miller
B. Tao

2.4. The Vapor-Phase Deposition of Clear Films of CdS

When a sample of CdS is raised to its sublimation temperature in an unreactive environment (i.e., rare gas, H_2, or a vacuum) it undergoes a dissociation into its elements, [2] thus: $CdS \rightarrow Cd^{o} + S^{o}$. Any further account of the product produced in a condensation process is primarily concerned with the recombination of the elements in the gas phase. This applies to both the vacuum sublimation followed by deposition upon a heated substrate, or to vapor transport in H_2 or gas mixtures. In both the vacuum system and the vapor transport, the condensed material is in some equilibrium with the elements in the gas phase.

The principle operative in this case can be demonstrated by observing bands of metallic Cd and Cd-rich CdS at the cold end of a furnace tube and bands of sulfur (emitting white light under U.V. irradiation) even farther from the heating element.

Should an excess of sulfur be introduced into the gas phase, a reaction occurs giving a clear yellow transparent film where previously there had been Cd^{o} and Cd^{o} rich CdS. A clear CdS can be obtained in this

way and is substantially the process that is operative in the vacuum deposition used by Bleha (see preceding section). In this case the deposition of sulfur seems to be restricted to the reaction product by maintaining a reduced pressure (10^{-6} Torr) and elevated temperature ($200^{\circ}C$) at the substrate. These temperatures and pressures inhibit the deposition of elemental sulfur, but encourage the reaction of the condensed Cd metal with the sulfur-rich gas phase. It is obvious, of course, that maintenance of a constant pressure differential is of primary importance in the preparation of a uniform product in such a system.

2.5. Vapor Transport in Gas

The method used for the preparation of the largest single-crystal CdS is a high-temperature ($1000^{\circ}C$), atmospheric-pressure synthesis in a tube in which a slow transport ($3ft^3/h$) is made of Cd° and S° into a hot region from boilers. [3] Under these conditions one obtains large single crystals. If CdS is used as the source material rather than Cd° and S°, relatively good single crystals are obtained, [4] although they are somewhat smaller. If a reduced pressure and lower temperature are applied the crystals become even smaller, but obvious hexagonal spines are visible. While single-crystal CdS is relatively transparent, the smaller crystals refract light so that only the thinnest films transmit an image. If the rate of gas flow in the sample chamber is increased, a uniform crystalline film is prepared to substantial thickness (10 μm)

and of crystallite size such as to permit the reading of ordinary printing in contact with the film. Film properties as related to gas flow are shown in the following table:

TABLE 2.1

Sample No.	Mole % In	H_2 Flow 10^{-4} moles/min	Ar Flow 10^{-4} moles/min	Mole Fraction H_2	Resistivity Ω/sq	Room T Luminescence
100	0.53	126	6	.95	15	Green
98	0.44	93	16	.85	39.	Green
106	0.56	69	17	.80	1.9×10^{6}	Red
101	0.57	93	37	.71	4.5×10^{6}	Red
108	0.00	93	16	.85	$>10^{10}$	none

All these data were obtained at a total pressure of 250 Torr. All samples were allowed to cool slowly.

From the data in Table 2.1 there seems to be a correlation of conductivity with mole fraction of H_2. H_2 reacts with S giving H_2S which passes through the tube leaving Cd^o to deposit in samples. [5] This mechanism agrees with our observation that the conductivity of CdS depends on the formation process. It was not possible to obtain resistivities of the order of magnitude of the samples prepared in high hydrogen concentration (15 - 30 Ω/sq) by baking high resistivity samples ($>10^5$Ω/sq) in pure H_2.

The complete mechanism cannot be explained in terms of Cd^{o} inclusion, however, since CdS prepared without indium doping (sample 108) is considerably more resistant than that with indium doping (98).

2.6. The Measurement of Conductivity

A number of attempts have been made to make conductivity measurements on films of CdS. In the estimation of the writer, none of these have been satisfactory. If one examines the work of Avis [6] and the figure from his publication (reproduced here as Fig. 2.3), the problem is obvious.

The problem presented by the need of ohmic contact alone brings the sample in contact with moisture. Moisture is a well-established variable in evaluating the conducting characteristics of CdS. [7,8]

The effect of water on the luminescence [8,9] of single crystals has been noted by various authors and has been observed in this laboratory to give a brilliant green photo luminescence to In doped CdS when entrained in the carrier gas (H_2). The effect of water may therefore not be totally undesirable, but its significance in conductivity measurements cannot be ignored.

The thermal cycles represented in Fig. 2.3 indicate that some mechanism of annealing out defects is operative. This is supported by Avis's report [6] of increased grain size on heat treatment. It is not clear that films prepared by vapor transport will behave in the same way as Avis's vacuum deposited films.

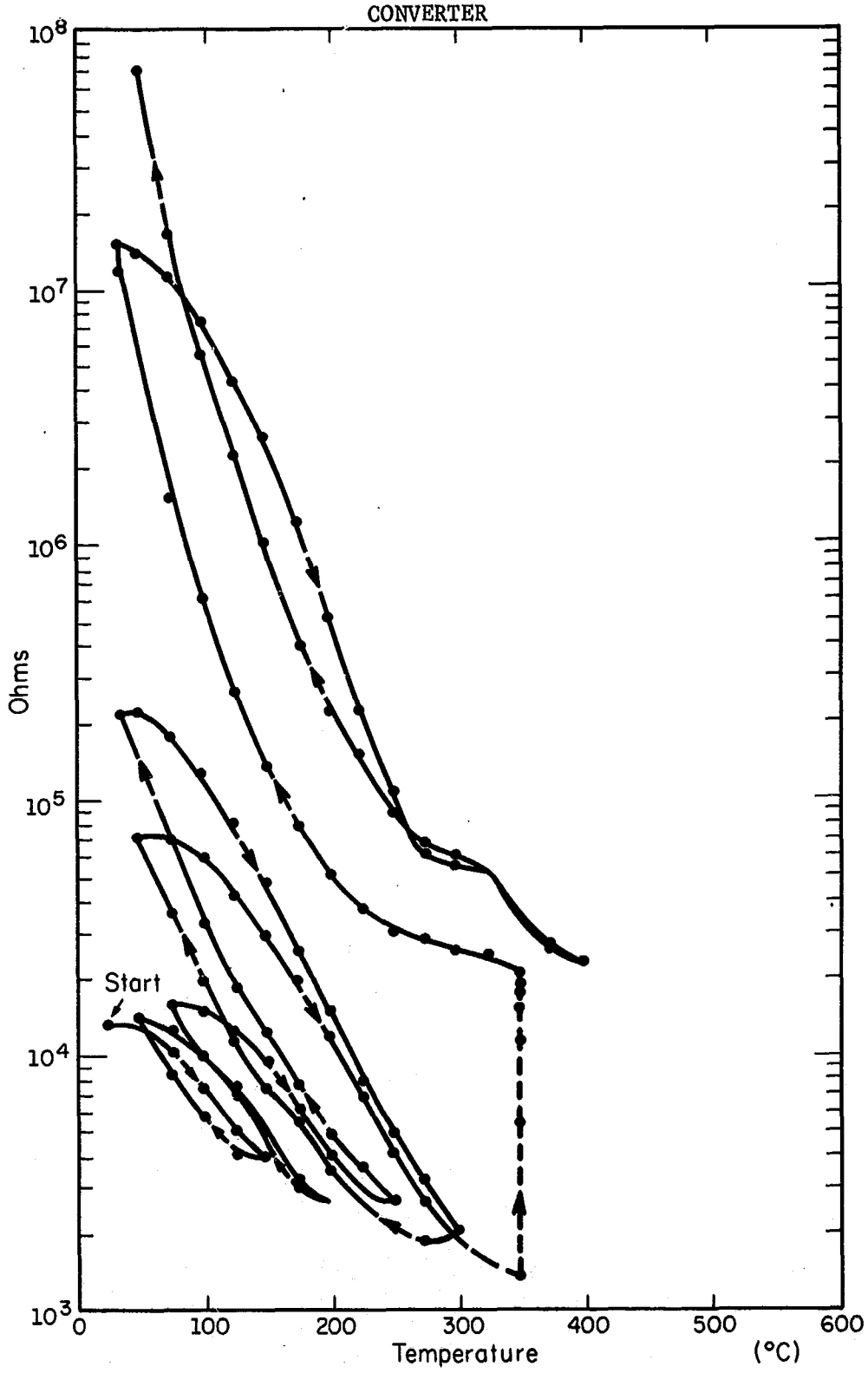


Fig. 2.3. Resistivity vs temperature for a CdS sample cycled to successively higher temperatures. LR-257

(OVERLEAF BLANK)

Samples 100 and 106 in Table 2.1 were heated to temperatures of 135°C (see Fig. 2.4), held at that temperature, and cooled. Resistivity measurements were made at various times in the process. The low-resistance sample (sample 100, 15 Ω/sq) seemed to be quite reversible. The high resistance sample is photo conducting and has a dark, room-temperature, resistivity of 1.9×10^6 Ω/sq. In both cases, contacts were made by evaporating indium on the Van Der Pauw samples.

Both samples have a decrease in resistance as the temperature is increased. This observation in consistent with an increase in the number of carriers as a function of temperature. The conduction mechanism in the case of the low-resistance sample seems to be reversible and linear over the temperature range (see Fig. 2.4).

A better interpretation of these data should be possible after further analysis of the samples.

A. Tulumello

2.7. Small Band-Gap Semiconducting Films

Emphasis has been placed on taking relative response and sensitivity measurements on photoconducting films of PbTe and PbS in the infrared. Relative response curves were made for PbTe cells that had undergone various types of post-treatment. In brief, it was found that PbTe films that have not been treated with oxygen have a decreased sensitivity at wavelengths greater than three microns. More sensitive measuring techniques are now being used and these data will need to be re-examined.

Absolute sensitivity measurements have been made on several films of PbTe and PbS. A calibrated black body was used as the light source, thereby providing a known intensity of radiation. Optical filters were used to filter the radiation to transmit a peak at 1.175 μm with a peak width of about 0.03 μm. The quantity D^* is computed as

$$D^* = (S/N)(\Delta f A)^{\frac{1}{2}}/W,$$

in which S is the signal voltage, N is the rms noise voltage, Δf is the bandwidth of the measuring equipment, A is the area of the cell, and W is the power hitting the cell.

A wave analyzer of bandwidth 3 Hz was used to measure the signal and noise voltages.

A summary of typical D^* measurements is given in Table 2.2.

TABLE 2.2

Cell	Treatment	D^* cm$Hz^{\frac{1}{2}}/W$	τ (sec)
PbTe 7-15-68	no oxygen	4.2×10^{7}	2×10^{-2}
PbTe 6-27-68	Oxygen treated	7.1×10^{6}	2×10^{-4}
PbS 7-16-68	Oxygen treated	1.2×10^{10}	4×10^{-4}

The cell with the highest value of D^* to date is a PbS, oxygen-treated film. The value of D^* for this cell is 1.2×10^{10} cm $Hz^{\frac{1}{2}}/W$ at a wavelength of 1.175 μm. This is comparable to values given for commercial InSb cells at this wavelength. An order of magnitude estimate of the cell response time τ was made by monitoring the ac photocurrent on an oscilloscope.

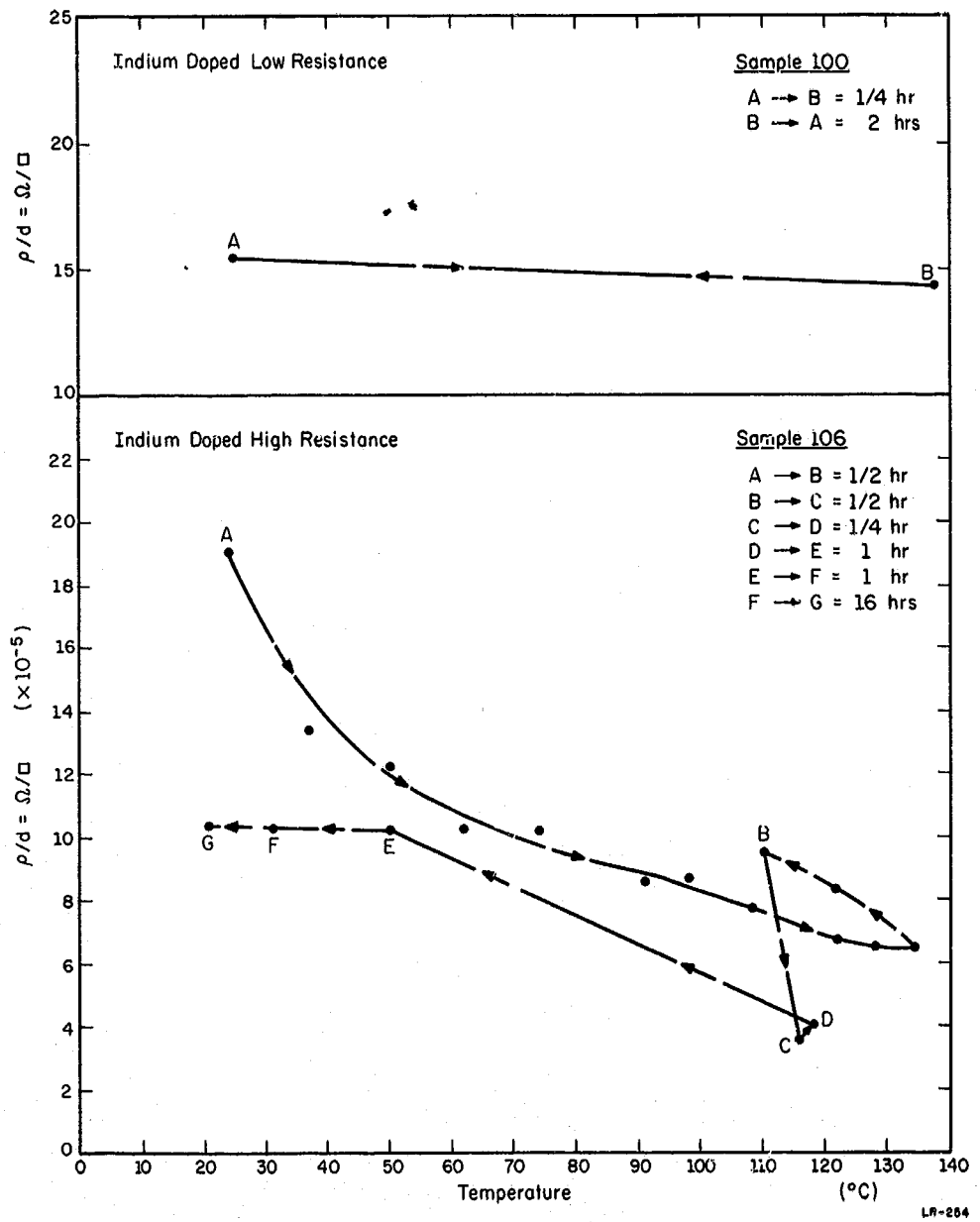


Fig. 2.4. Resistivity vs temperature for an indium-doped CdS sample.

(OVERLEAF BLANK)

It should be noted that oxygen had little effect on the PbTe films with the exception of lowering the response times. This may be due to the fact that the PbTe source material is already p doped.

Corrections for light losses in the optical filter and in the dewar window were made in the computation of D^*. The per cent error in the D^* values given is approximately 50 per cent or less, the largest source of error being the signal and noise voltage readings.

J. E. Robinson

2.8. Device Fabrication and Testing

Work has continued on the single-crystal electroluminescent devices comprising a CdS-Te heterojunction. Many of these units have been fabricated and tests have been run to categorize their behavior. Luminescence becomes visible in these diodes at roughly 10 volts applied bias, the CdS side being positive. The emitted light is due to edge luminescence in the CdS, its spectrum peaking at approximately 0.515 μm and its intensity is exponential in applied voltage over a several volt range.

Work is progressing on an infrared-sensitive switch formed of a Ge-Te interface. Further tests are being made to clarify the theory of operation and to optimize its behavior.

An electroluminescent unit and a switch unit have been series connected and tested in liquid nitrogen. The former gave off green light

when the switch unit was illuminated with IR. With this experiment as a basis, composite units have been fabricated consisting of In-CdS-Te-Ge-Pb:Sn:Sb. Figure 2.5 shows one such unit that has been sectioned and photographed in the scanning electron microscope. None of the composite units fabricated to date act properly as IR converters. Work is continuing to solve the fabrication problems.

G. E. Anner
B. Tao

2.9. Tunneling

2.9.1. Experimental

The next portion of this period involved work on the tunneling properties of germanium nitride, a possible insulator for use in the infrared-visible light converter. It is believed that germanium nitride decomposes at approximately 600°C; thus it is of little use to transistor technology. As a result, germanium nitride properties have not been investigated to the same extent as have other insulators, such as silicon nitride.

An experiment was designed to test the suitability of germanium nitride as a tunneling barrier. The preliminary results of this experiment indicate that tunneling barriers are easily formed on air-cleaved germanium by plasma anodization in a nitrogen ambient. It should be noted that these barriers may not actually be germanium nitride, as no test of either the composition or stoichiometry of the insulator was made.

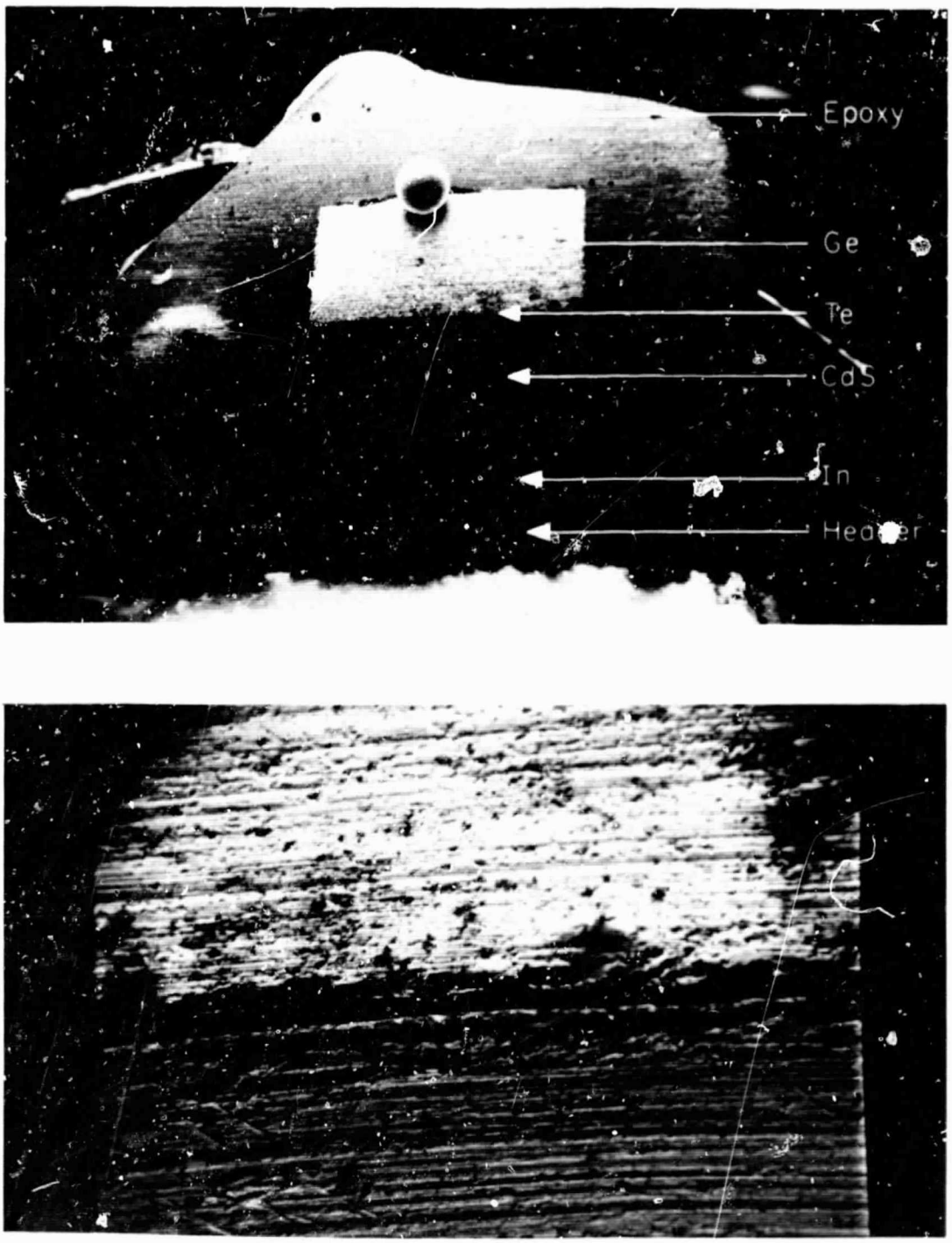


Fig. 2.5. Scanning electron microscope photograph of single-crystal electroluminescent device. The upper photograph shows the sectioned composite unit with 4.5 V applied. The small sphere is ohmic-contact alloy that had melted out when bias was applied with no cooling. The lower photograph shows an enlarged view of the Ge-Te-CdS region.

(OVERLEAF BLANK)

The observation of the lead superconducting energy gap at $4.2^{\circ}K$, in structures of the type Pb-insulator-Ge (0.001 Ω cm, n-type), indicates that the plasma-anodized layer was compatible with superconducting tunneling. The room-temperature, zero-bias resistivity of a junction without the anodized layer was compared to the resistivity of one formed by anodization. The anodized sample was several orders of magnitude more resistive than the unanodized one. This, together with the observation of the superconducting energy gap, indicates conclusively that insulating films of tunneling thickness can be grown on a germanium surface by plasma anodization.

2.9.2. Theoretical

Also during this period, it was decided that a simple method of determining the barrier heights of the insulator in tunneling structures was needed. Therefore a theoretical study was initiated in which a computer program for tunneling current written by Dr. L. C. Davis of the University of Illinois was used. This program calculated the current J vs bias V for an M-I-M structure at various temperatures. It was then possible to construct a variety of functions of J, dJ/dV and d^2J/dV^2 and look at their behavior with bias and temperature T. Certain functions, such as

$$\hat{J}(V) = [J_1(V)-J_2(V)]/J_2(V)$$

in which the subscripts refer to the temperatures, $T_1 > T_2$, show definite peaks at bias values near the barrier heights. The complete analysis of the calculations should show which function most clearly indicates the barrier heights of the insulator for a given range of temperature and barrier heights.

J. T. Jacobs
S. Depp

2.10. Size Effects

2.10.1. Experimental

The results reported in R-364 "Size Effects in the Temperature Variation of Electrical Conductivity of Epitaxial Gold Films" proved the worth of the temperature-variation technique for studying electrical size effects. The equipment used previously has been improved considerably by construction of a new sequential programmer for automatic data recording on punched tape. Unfortunately, when it was first used, a calibrated germanium low-temperature thermometer failed, causing a delay. A replacement has now been obtained and new measurements are in progress.

The computer programs for data reduction have also beem improved.

J. T. Jacobs
R. C. Birtcher

2.10.2. Theoretical

The size effects study of electrical conduction in thin films led to interest in the abnormally high specularity of the surface of bismuth

films. Dr. C. B. Duke of the University of Illinois and General Electric suggested that a surface space-charge region might cause this high specularity as well as bismuth's anomalous temperature dependence for optical absorption. A theoretical calculation was then begun to determine the character of the space charge at the surface of a simple one-carrier solid having some of the parameters of bismuth.

It was hoped that the temperature dependence of the charge distribution would show that the carriers move closer to the surface as the temperature increases, therefore changing the optical absorption. The calculations showed that for constant surface field, this movement did indeed occur. Further investigation into the program of optical absorption in degenerate semiconductors and in semimetals showed, however, that interband transitions would likely overshadow this mechanism. Therefore, work on this problem was discontinued.

2.11. References

1. J. Fassbender, Z. Physik 145, 301 (1956).

2. M. Aven and J. S. Prenner, Physics and Chemistry of II-VI Compounds (North Holland Publishing Company 1967) p. 78.

3. R. H. Fahrig, Electrochemical Technology 1, 362 (1963).

4. See, for example, N. A. Komrons "Cadmium Sulfide, A History of Semiconductor Research at the Aerospace Research Laboratories," Superintendant of Documents, U. S. Government Printing Office (Washington, D. C. 1964).

5. F. A. Kröger, The Chemistry of Imperfect Crystals, North Holland Pub. Co. (1964) p. 182.

6. G. G. Avis, et al., U. S. Department of Commerce Report No. AD622 695 Unclassified (July 20, 1965).

7. P. Faeth, J. Electrochem. Soc. 114, 511 (1967).

8. C. E. Bleil and W. A. Albers, Surface Sci. 2, 307 (1964).

9. P. Faeth, private communication.

M. Raether
H. Böhmer
D. Bollinger
J. Chang
K. Evans
D. Fenneman
R. Holden
R. Hosken
D. Kraybill
D. Stewe

3.1. Observation of Surface Modes Excited by an Electron Beam in a Plasma

During measurements of the spatial development of the electron beam-plasma instability we find resonances in the unstable wave spectrum that do not fit into the ordinary picture of beam-plasma instabilities.

It can be shown that during the interaction of a cold electron beam with a cold, collisionless, and infinite plasma, waves become unstable whose optimal frequency ω and growth rate γ are given by [1]

$$\mathrm{Re}(\omega) = \omega_p[1-\tfrac{1}{2}(\alpha/2)^{2/3}]$$

$$\gamma = \mathrm{Im}(\omega) = \omega_p\tfrac{1}{2}\sqrt{3}(\alpha/2)^{1/3}$$

where ω_p is the plasma frequency and α the ratio of beam to plasma density. For most experiments there obtains $\alpha \ll 1$. In a semi-infinite plasma, the waves grow simultaneously in time and space, a problem that has not yet been solved satisfactorily. Following Kadomtsev [2] and Fainberg and Shapiro [3] one can, under certain conditions, relate the temporal growth rate γ with the spatial growth rate σ via

$$\sigma = \gamma/v_g$$

[†]This work was supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB07-67-C-0199.

where v_g, the group velocity of the unstable waves, is given by $v_g = 2v_B/3$, with v_B being the beam velocity. From these results, one concludes that the unstable waves grow exponentially in space with a frequency close to the plasma frequency.

To study the spatial development of waves in a beam-plasma system, an experimental system described in Fig. 3.1 is used. A pulsed electron beam, 0.1 to 1 ampere, 18 kV and 1 cm diameter is injected into a decaying Neon afterglow plasma with an initial density of 10^{14} cm^{-3} and an average decay constant of 150 μsec. The Pyrex plasma tube is 4 cm in diameter and 40 cm long. No external magnetic field is applied to the system. The experiment is performed with a repetition frequency of 15 Hz. The electron beam can be exposed to different plasma densities by changing the time delay between the plasma and the beam pulse. A microwave receiver (26.5 to 40 GHz, 11 dB noise figure) receives the microwave power radiated by the instability. The output of the receiver is detected in a synchronous fashion by a lock-in amplifier. The position of the receiver and the detection horn can be varied along the plasma tube. Using a microwave lens, the spatial resolution is approximately 1 cm. The polarization of the receiving horn is such, that the E-vector is parallel to the electron beam. For each position of the receiver the radiated power is recorded as a function of plasma density.

Figure 3.2 shows some samples of such recordings for several positions of the receiver. Apart from higher harmonics of the plasma

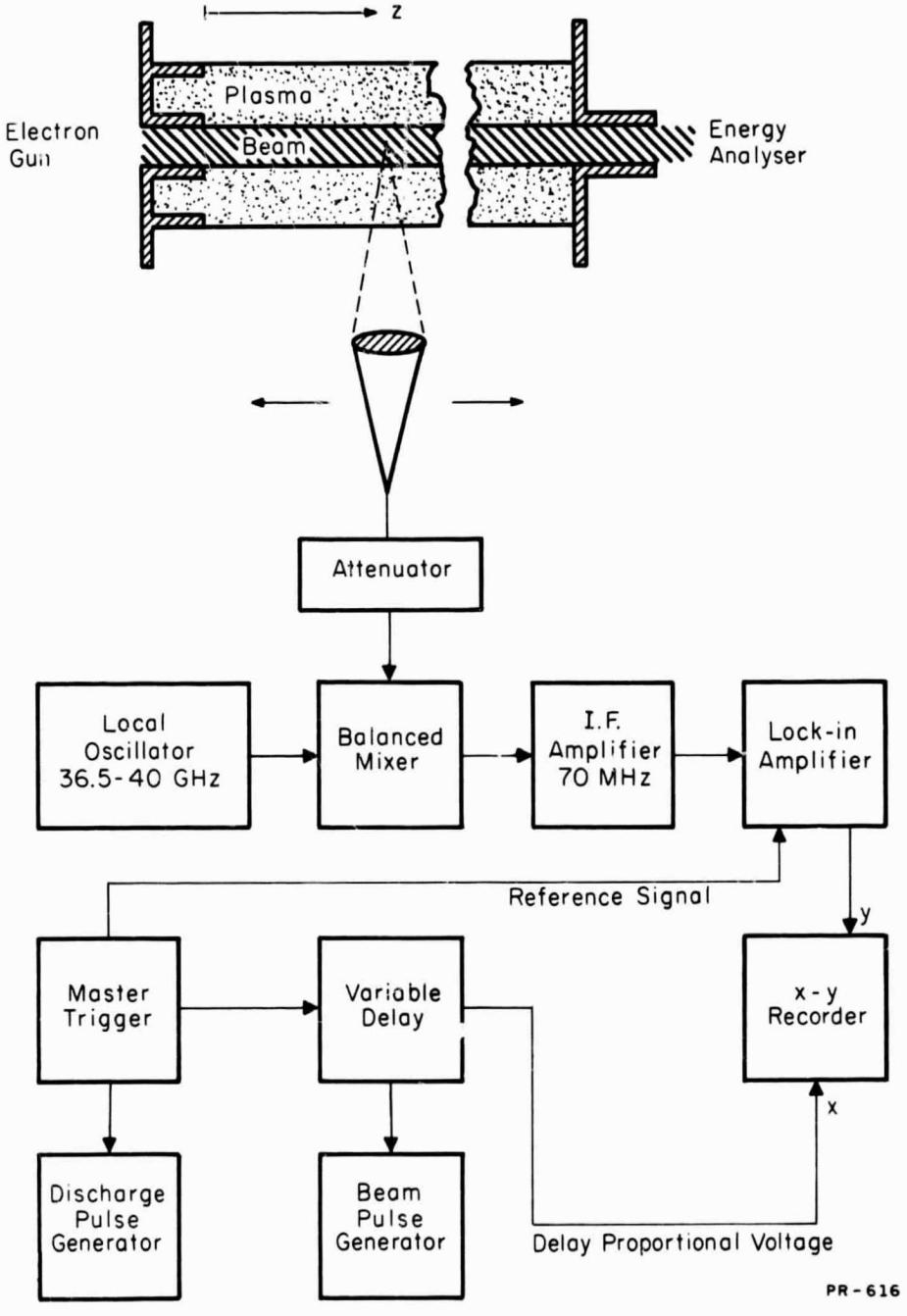


Fig. 3.1. Schematic of the experiment.

(OVERLEAF BLANK)

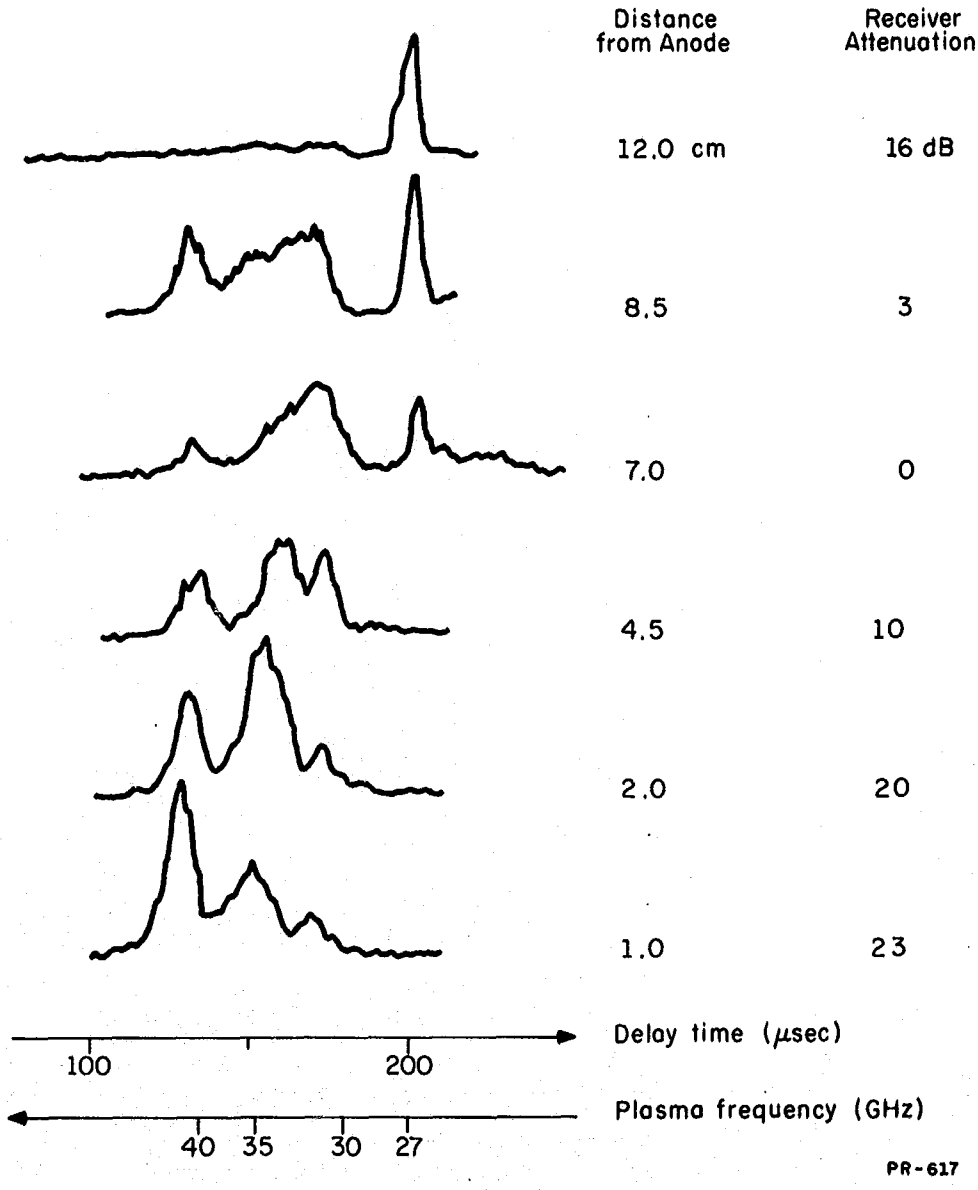


Fig. 3.2. Received power versus delay time and plasma frequency for several positions of the receiver. Receiver frequency: 26.5 GHz, electron beam: 170 mA, 18 kV; pressure: 54 Torr (Neon).

(OVERLEAF BLANK)

frequency which occur at much longer delay times and are not shown here, two frequency groups are observed. At long distances from the point where the beam enters the plasma, the signal occurs when the receiver frequency is equal to the plasma frequency. This is the beam-plasma instability referred to in the introduction. It exhibits the expected exponential increase with distance (Fig. 3.3).

At smaller distances, a signal that consists of several peaks and becomes stronger with decreasing distance is received at delay times where the receiver frequency is much smaller than the plasma frequency. The amplitude of two such resonances are plotted as a function of distance in Fig. 3.3. These curves are corrected for the angular response function of the receiving horns. From the magnitude of the radiated power we conclude that, for example, at z = 1 cm, the fluctuation amplitude of the plasma was increased by a factor of 10^5 to 10^6 above the thermal equilibrium value.

It should be noted that the same resonances, although with a much smaller amplitude, are observed at the opposite end of the discharge tube, i.e., at the point where the beam leaves the plasma.

In starting the discussion about the origin of these modes, we have to note that the beam was not premodulated in any way. Furthermore, we have to rule out the "virtual cathode" [4] as the source of this radiation because, using Nezlin's [4] criterion for the onset of this effect in charge compensated beams, our electron beam density is too

small by a factor of 10^3. Moreover, measurements with an electrostatic deflection type energy analyzer show that the change in the energy distribution function of the electron beam due to the excitation of these waves was less than 0.3%. For these measurements, the interaction length was decreased to 6 cm, so that only the low-frequency waves were excited.

Since we perform the experiment in a decaying plasma, a radial as well as an axial density gradient will be present, the latter only close to the ends of the discharge tube. The beam has to traverse these axial density gradients in entering and leaving the plasma. We would like to suggest that the oscillations observed at small z are the eigenmodes of the inhomogeneous plasma boundary driven into instability by the electron beam.

It is well known that in an inhomogeneous plasma, surface waves can exist. These modes have been investigated in detail in a number of publications. [5] In a cylindrical plasma geometry these modes can be excited by radial transmission of microwaves (Tonks-Dattner resonances). They present themselves as a series of resonances in the absorbed or reflected power.

In a semi-infinite plasma, bounded by a density gradient, one can look at the surface modes as standing waves reflected at the vessel wall on the one side and at the point where the frequency of these waves is equal to the plasma frequency on the other side. Beyond this point,

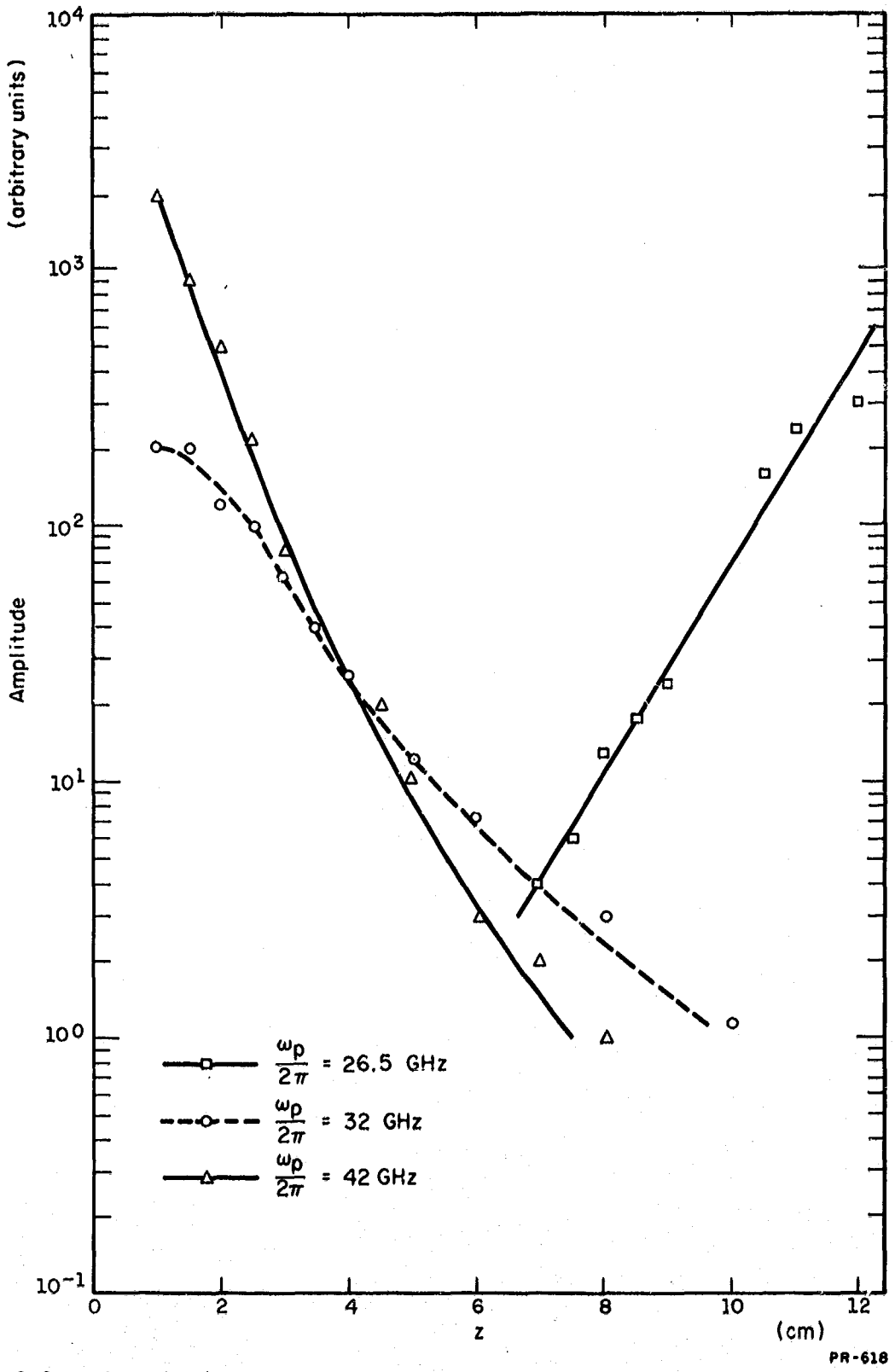


Fig. 3.3. Received power versus distance for the instability in the homogeneous plasma ($\omega_p/2\pi$ = 26.5 GHz) and for two resonances ($\omega_p/2\pi$ = 32 GHz and $\omega_p/2\pi$ = 42 GHz). The experimental conditions are the same as in Fig. 3.2.

(OVERLEAF BLANK)

where $\omega_p > \omega$, the waves will decay exponentially. Hoh [6] treats this problem assuming a density profile given by $n = n_L Z/L$. He finds for the mode frequencies for a non-magnetized plasma

$$\omega^2/\omega_{pL}^2 = [(3\pi\lambda_D/2L)(m+3/4)]^{2/3}, \quad m = 0,1,2,\ldots,$$

where ω_{pL} and n_L are the plasma frequency and density at the point $z = L$. The quantity λ_D is the Debye length. Since $\lambda_D/L \ll 1$ for most experiments, the frequency of these modes will be smaller than ω_{pL}. Other density profiles can also be treated and give qualitatively similar results. The observed frequencies do not coincide with the resonances calculated by Hoh for any value of λ_D/L, nor do they agree with those calculated from the more realistic profile $n(z) = n_o(1-\exp(-z/L))$. The reason for this is probably that the beam modifies the eigenfrequencies in a more drastic fashion than in the case of a homogeneous plasma. This would not be too surprising, because the intuitive picture of a standing wave between the boundary and the plane where $\omega = \omega_p$ will lose its validity in the presence of a beam since the beam-plasma system can support waves of frequencies less than ω_p that are located on the unstable branch of the dispersion relation.

In the presence of an electron beam one may expect, in analogy to the infinite beam-plasma system, that instabilities occur at frequencies close to the eigenmodes. Since these waves exist only in the region of finite density gradient, they will not exhibit a convective

growth; in other words, their group velocity will be zero. The development of such an instability closely resembles a case treated by Fainberg and Shapiro [3] who investigated the temporal and spatial development of an instability under the hypothetical condition of zero group velocity of the unstable waves. The plasma was assumed to be semi-infinite. They found that the system is unstable and that the oscillations accumulate in a narrow layer close to the plasma boundary in a similar fashion as observed in our measurements. In their treatment, of course, the waves become unstable at the plasma frequency, whereas in the presence of a density gradient the eigenmodes that can be excited are the surface modes.

H. Böhmer
M. Raether

3.2. Generation of Smeared-Out Electron Beams for Experiments on Weak Plasma Turbulence

The theory of weak plasma turbulence [7,8] can make very detailed predictions about the nonlinear interactions of a "smeared out" charged particle beam with a plasma. Although an impressive amount of theoretical work exists in this area, relevant experiments are essentially lacking. This is largely due to the fact that the theory of weak turbulence imposes rather stringent conditions on the necessary velocity spread of, say, an electron beam, which have been difficult, if not impossible to implement experimentally.

It is generally considered as a sufficient criterion for the applicability of the theory of weak turbulance that the velocity spread Δv is large compared to γ/k_{opt}, where γ is the linear growth rate and k_{opt} the wavenumber for which γ becomes a maximum. [7,9] For a beam distribution function of the form of a displaced maxwellian

$$f_B \propto \exp[-(v-v_o)^2/2\Delta v^2]$$

the growth rate is given by

$$\gamma = 0.38(\omega_B^2/\omega_p)(v_o/\Delta v)^2$$

and $k_{opt} = \omega_p/v_o$. Here ω_B and ω_p are the plasma frequencies of beam and plasma electrons respectively. With these values the conditions on the velocity spread can be written:

$$(\Delta v/v_o) \gg 0.725(\omega_B/\omega_p)^{2/3}$$

Inasmuch as the transition from a "cold" beam to a "hot" beam takes place when [10]

$$\Delta v/v_o \approx (\omega_B/\omega_p)^{2/3} ,$$

it may be expected that the condition for weak turbulance already obtains when $\Delta v/v_o$ is of the order of $(\omega_B/\omega_p)^{2/3}$. In a typical experimental situation, the ratio of beam to plasma density ω_B^2/ω_p^2 may range from

10^{-6} to 10^{-3}, which would require velocity spreads of the order 1 to 10%. This value should be contrasted to the velocity spread of a beam from a hot cathode, which for the energies considered here is of the order 10^{-2} to 10^{-3}%.

We have produced beams with velocity spreads of a few percent by passing a pulsed electron beam through aluminum foils 1000 - 4000 Å thick. The foils were produced by vacuum deposition of aluminum onto glass slides coated with an aerosol-sugar substrate. The thickness of the foils was measured by a Deposit Thickness Monitor* with an accuracy of $\pm$ 10 A^{o}. After the deposition, the substrate was dissolved in water and the foils were mounted onto a 90%-transparent, stainless-steel grid of 1-mm mesh size. A pulsed electron beam of 17.6 kV, 1 A and 10 μsec duration was passed through these foils at a repetition frequency of 15 Hz and its distribution function measured with a deflection-type electrostatic energy analyzer of 0.3% resolution.

Figure 3.4 shows a typical example of the so-obtained distribution functions. The geometry was such that only electrons scattered into a forward angle of less than $\pm$ 1^{o} were measured by the energy analyzer. The distribution functions are not displaced maxwellians; we therefore choose the rms value of Δv, denoted by $\langle \Delta v \rangle$, as the parameter characterizing

*Model DTM-3, Sloan Instrument Corp.

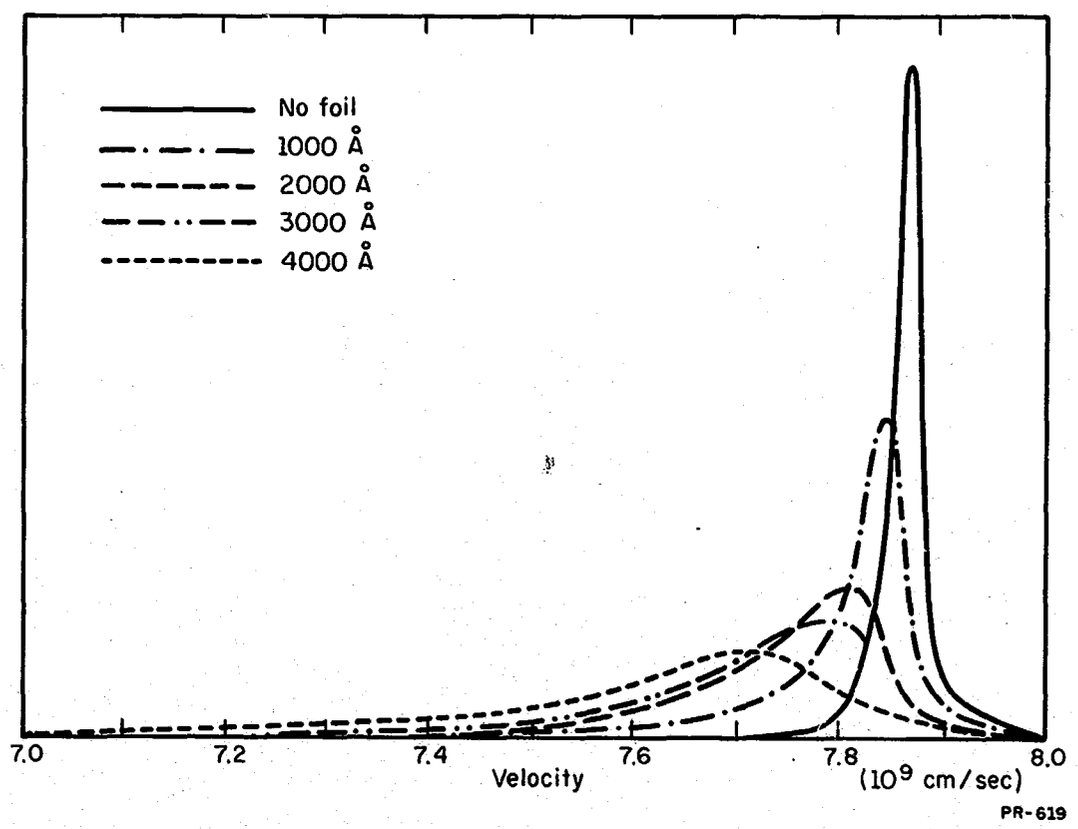


Fig. 3.4. Distribution function of a 17.6 kV, 1A pulsed electron beam after passing aluminum foils of different thickness. The width of the distribution function of the original beam is determined by the resolution of the electrostatic-energy analyzer.

(OVERLEAF BLANK)

the velocity spread. $\langle \Delta v \rangle$ is plotted in Fig. 3.5 together with the measured transmission. After running the experiment for several hours at pulse current densities of 1 A/cm^2, the foils did not show any noticeable deterioration.

The data show that it is possible to generate electron beams of good intensity with velocity spreads of a magnitude that makes it feasible to perform experiments on weak turbulance. It should be pointed out that, for beam plasma interactions in magnetic fields, the situation is even more favorable because the low energy electrons that are scattered out of the beam will be confined by the magnetic field. One therefore would expect much larger velocity spreads and better transmission.

Finally, it is worth mentioning that beam distribution functions with similar velocity spread are generated by the beam-plasma instability itself. Figure 3.6 shows an example of an originally monoenergetic beam of 17 kV, 500 mA, and 10 μsec duration after interacting with a 30 cm long plasma column for different plasma densities. This method has, however, several disadvantages. The velocity spread is a sensitive function of not only the plasma density but also of the beam current, beam diameter, interaction length, and other factors that depend on the particular experimental situation. Moreover the beam is modulated at the plasma frequency corresponding to the plasma density at which the interaction takes place. We have measured this modulation by injecting a known

fraction of the beam into an X-band waveguide and measuring the microwave power at 8.4 GHz. The electric field on the beam was determined to be of the order of 200V/cm. Although this indicates that most of the velocity spread is indeed random, the rather large premodulation may present problems in certain experimental situations. Such a modulation is of course absent when foils are used since the velocity spread is only due to elastic and inelastic collisions in the metal. [11]

In conclusion, it appears that the preparation of electron beams with smeared-out distribution functions by passing the beam through thin foils might be a useful technique to study weak turbulance effects in plasmas. Experiments in this direction are in progress.

We are much indebted to Dr. F. Steinrisser for help and advice in the preparation of the foils.

J. Chang
H. Böhmer
M. Raether

3.3. Experimental Investigation of Nonlinear Interactions of Drift Waves [12]

A single-ended Q machine was built and collisionless cesium and potassium plasmas are investigated.

It is demonstrated that the observed frequency of drift waves in a Q machine can be varied by changing the radial electric field *via* a voltage applied to the conducting vacuum envelope. Thus there is additional verification that drift waves are subject to azimuthal doppler shift.

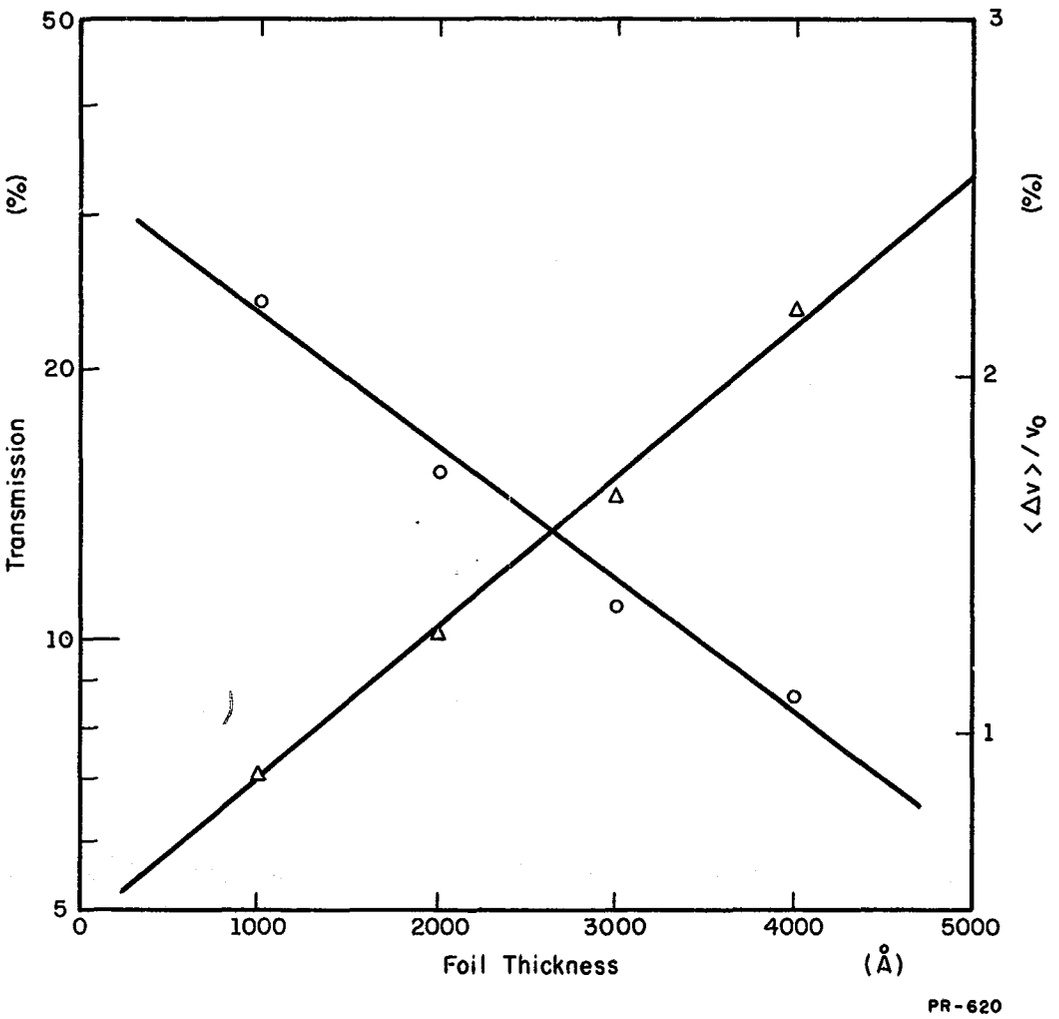


Fig. 3.5. Root-mean-square velocity spread and transmission as a function of foil thickness. The circles refer to the transmission; the triangles to the velocity spread.

(OVERLEAF BLANK)

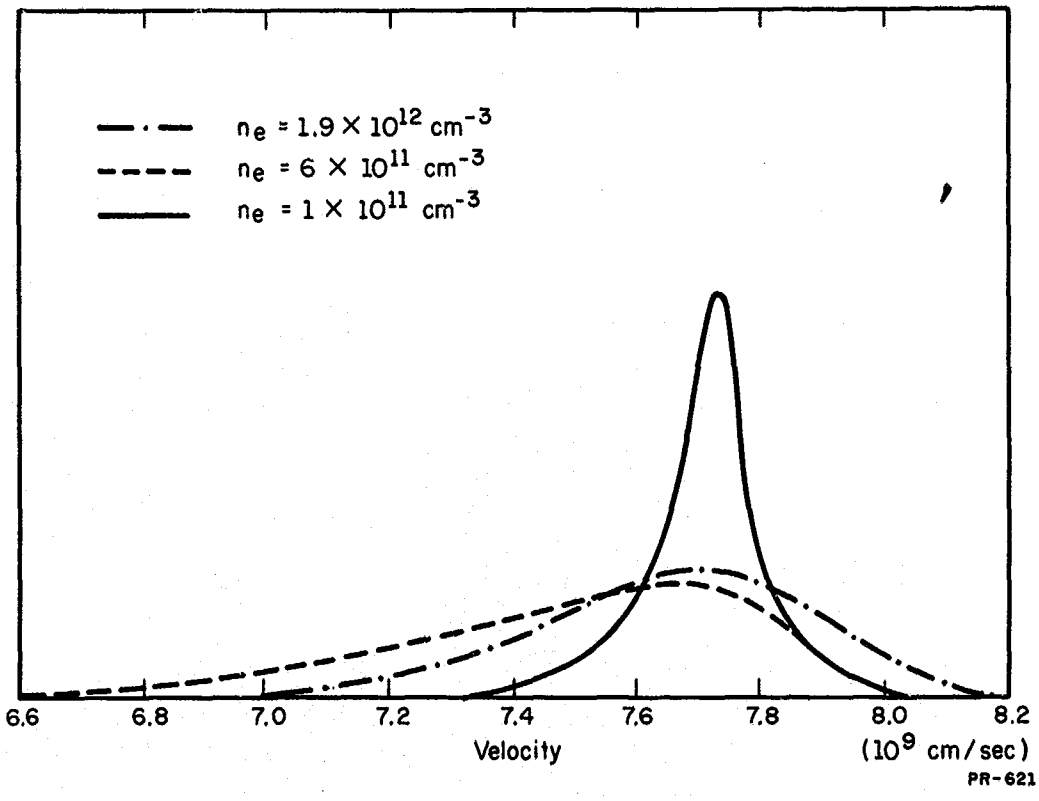


Fig. 3.6. Example of an electron beam that acquired a velocity spread by interaction with a plasma at different plasma densities. The beam parameters were 17 kV, 500 mA, 10 μsec duration, the interaction length was 30 cm.

(OVERLEAF BLANK)

The nonlinear interaction of drift waves in a Q machine generates new waves at the sum and difference frequencies. In particular, two waves of the same order of magnitude can create a new wave that is only an order of magnitude smaller in amplitude. Lashinsky's [13] concept of the synchronization of a higher-order mode to the nonlinearly generated signal appearing at integer multiples of the fundamental is put on a sounder basis now that it is demonstrated that the mixing products have a finite amplitude. The higher-frequency product waves will have less amplitude. However, once the second mode is locked to twice the fundamental, it, in turn, will mix with the fundamental to create a signal at the third harmonic, and so on. As we have shown in general, if all the waves occur equally spaced in frequency, then more energy is available for the discrete modes at higher frequencies.

When the mixing products are not equally spaced, the discrete waves at higher frequencies are less pronounced and the energy appears in the continuum. This may compare to the ordinary situation in a Q machine where "periodic pulling" is believed to create sidebands that mix with the normal modes to create the continuum. The presence of two sets of waves in the present experiment allows an observation of the continuum as the ratio of the frequency of the waves is changed. It is demonstrated that the continuum suffers a loss in amplitude when the mixing products are made commensurate in frequency.

The spectral dependence of the continuum created by drift waves falls off as $\omega^{-3/2}$. This measurement is made in a plasma exhibiting the two modes and their mixing products. However, over most of the volume of the plasma, the ordinary drift wave is the dominant feature, and this experimental value is determined from a volume integral. The dispersion relation for drift waves is approximately linear in the rest frame of the plasma. The doppler shift makes the dispersion relation more linear in the laboratory so that one can write $I_k \propto k^{-3/2}$ for drift waves.

R. Hosken
M. Raether

3.4. References

1. V. S. Imshennik and Yu. I. Morosov, Sov. Phys.-Tech. Phys. 6, 464, 1961.

2. B. B. Kadomtsev, Plasma Turbulence, Acad. Press, London, 1965.

3. Ya. B. Fainberg and V. D. Shapiro, Sov. Phys. JETP 20, 937, 1965.

4. M. V. Nezlin and A. M. Solintsev, Sov. Phys. JETP 22, 949, 1966.

5. A. Dattner, Phys. Rev. Letters 10, 205, 1963;

   T. V. Parker, J. C. Nickel and R. Gould, Phys. Fluids 7, 1489, 1964;

   P. Weissglas, J. Nucl. Energy, Pt. C6, 251, 1964;

   F. W. Crawford, J. Appl. Phys. 35, 1365, 1964.

6. F. C. Hoh, Phys. Rev. 133, A1016, 1964.

7. W. E. Drummond and D. Pines, Nuclear Fusion, 1962 Supplement, p. 1049;

   W. E. Drummond and D. Pines, Annals of Physics, 28, 478, 1964.

8. A. A. Vedenov, E. P. Velikhov and R. Z. Sagdeev, Nuclear Fusion, 1962 Supplement, p. 465.

9. V. D. Shapiro, Sov. Phys. JETP 17, 416, 1963.

10. V. S. Imshennik and Yu. I. Morozov, Sov. Phys. - Tech. Phys. 6, 464, 1961.

11. L. Landau, J. Phys. USSR, 8, 201, 1966.

    R. D. Birkhoff, "The Passage of Fast Electrons Through Matter," Handbuch der Physik, Vol. XXXIV, 1958.

12. Robert W. Hosken, Thesis, Univ. of Ill. 1968.

13. H. Lashinsky, Dynamics of Fluids and Plasmas, Academic Press, Inc. (1966), p. 473.

(OVERLEAF BLANK)

4. RAREFIED GAS DYNAMICS†

B. L. Hicks
B. J. Reilly
K. R. Roth
W. H. Chu
S. M. Yen
A. Nordsieck*

4.1. The Boltzmann Computer Program

To replace laborious hand calculations used earlier, a computer program was developed for estimating random and systematic errors of the Monte-Carlo evaluation of the collision integral near thermal equilibrium. Iterated chi-square tests were made of the random-number generator used in the Monte-Carlo calculations and gave no evidence of nonrandomness of the pseudorandom numbers that are being generated. A new FORTRAN program allows computation of the error in numerical evaluation of the derivatives df/dn which is an essential part of our iterative solution of the Boltzmann equation. Mr. Chu has started to rewrite our Monte-Carlo program for ILLIAC IV. The first stage is writing the program in Algol language for a Burroughs computer which can be used to simulate ILLIAC IV. The first Algol program, one for calculating analytical values of a particular function a_n over velocity space, has been finished.

4.2. The Boltzmann Shock Wave

As reported earlier, solutions of the Boltzmann equation have been obtained for shock waves in gases of elastic spheres for a range of

†Supported by the Office of Naval Research under Contract No. ONR N00014-67-A-0305-0001 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

*Consultant.

Mach numbers from 1.1 to 10. In the past six months, the range was extended down to 1.05 and all of the solutions were refined. Selected results were prepared for the Sixth International Symposium on Rarefied Gas Dynamics at MIT. Space limitations in the publication of this paper suggest that we should publish much more fully elsewhere. Analysis of other physically interesting characteristics will proceed in the next quarter.

Means were developed for making careful comparisons of the Monte-Carlo solutions of the Boltzmann equation with several alternative expressions of the Navier-Stokes solutions for weak shock waves. The Maxwell-Boltzmann corrections described in the previous Progress Report yield values of the shock thickness, even for low Mach numbers like 1.2, which agree very well with Navier-Stokes values. A new series of error studies was also completed and showed in some detail the large effect of this Maxwell-Boltzmann correction in decreasing the Monte-Carlo errors.

Roth obtained the accurate six-moment solution of a generalized Mott-Smith shock for $M_1 = 2$ to 10 and for both rigid elastic spheres and Maxwellian molecules. The distribution function, its moments, and the functionals of the moments having physical significance have been compared with the corresponding Mott-Smith and Monte-Carlo results for some selected Mach numbers.

For many years, both theoreticians and experimentalists have compared the reciprocal shock thickness and the density profile calculated from one of three possible Mott-Smith solutions (with the u_x^2 moment equation) with their results. Since the six-moment solution satisfies three non-invariant moment equations rather than one, it is less arbitrary than the

Mott-Smith solution; therefore, the six-moment solution will be of interest to those people wishing to compare their results with the Mott-Smith model which is bimodal.

4.3. Heat Transfer in Rarefied Gases

Computations have now been made for complete energy accommodation for the following wide ranges of parameters: (1) Kn: 0.1 to 100, (2) M_T, a gradient parameter: 0.03529 to 35.29, and (3) T_1/T_2: 0.1 to 0.9. The solution was also obtained for the following values of parameters to match those of one of the experiments by Teagan and Springer: the energy accommodation coefficient α = 0.826, Kn = 0.631, and T_1/T_2 = 0.783. In order to permit appropriate comparison with our Monte-Carlo results, certain calculations have also been repeated for the Krook equation and the four-moment Boltzmann transport equations. On the basis of these calculations several studies have been made: (1) a study of the convergence of our numerical method and the uniqueness of the Monte-Carlo solution; (2) detailed analysis of our heat-transfer results; and (3) comparison with the corresponding Krook and four-moment results. The fractional error of $df/d\tau$, where τ is the integration variable, $Kn\sqrt{(2t/\pi)}\ n$, has been measured and the accuracy of four-moment and five-moment solutions of the heat-transfer problem have been tested and compared with the Monte-Carlo solutions.

The determination of detailed characteristics in a Knudsen layer in rarefied-gas dynamics problems is as essential as that of studying the boundary layer for aerodynamic problems in continuum-flow regime. One of

the significant results obtained is detailed evidence for the Knudsen layer, for certain Knudsen numbers and temperature ratios. Both temperature and density profiles exhibit nonlinearities near each boundary, for example, for Kn = 1 and 0.5, $T_1/T_2 = 0.7$, and thus show the presence of the Knudsen layer for these two values of Kn. The corresponding profiles for the Krook and four-moment solutions are practically linear and, therefore, do not show such a nonequilibrium behavior. New studies of this nonequilibrium behavior (Boltzmann) in the Knudsen layer have included calculations for several parameters and for a larger number of stations near both plates bounding the rarefied gas. Some of these results were included in the paper presented at the MIT Gas-Dynamics Symposium. Calculation of many functionals of the collision integral also aid this detailed study of the Knudsen layer.

In comparing the density profile calculated from our solution of the Boltzmann equation with that determined from the experiment by Teagan and Springer mentioned above, there is good agreement between the two results in the linear part of the profile but not in the nonlinear region near each boundary.

5. HIGH-VOLTAGE BREAKDOWN[†]

E. M. Lyman
D. Alpert
S. Y. Ettinger

T. C. Casale

T. Duchamp
J. Hansen
C. Stewart

5.1. Study of Gas-Conditioning Effects on Broad Area Electrodes

5.1.1. Experimental Results

In the previous report, [1] preliminary results of gas-conditioning tungsten and copper electrodes were described. In the present experiments utilizing the same apparatus as described by Alpert [2] et al., the anode was cooled to 80°K, the cathode kept at room temperature and the pressure, following pump down after the gas conditioning, was about 2×10^{-10} Torr. In a typical surface-conditioning experiment, argon gas is introduced into the system at a pressure of about 2×10^{-3} Torr. The voltage is gradually increased, keeping a steady field-emission current of about 50 μA. During the process, ions formed in the electron beams emitted from the tips of sharp cathode protrusions fall back on the protrusions, blunting or otherwise altering them by selective sputtering, thus suppressing the emission current and permitting an increase in the voltage. [3] After a few hours of conditioning, the vacuum breakdown strength, always measured after pump down, reaches an asymptotic limit of 3 to 5 times the initial value.

[†]This work was supported in part by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

During conditioning, a qualitative change occurs in the visible transition radiation pattern on the anode. The pattern due to "natural whiskers" present after a vacuum breakdown, or after the electrodes have been touched together, consists of a few intense spots indicating that the emission current arises chiefly from a small number of emitters on the cathode. On the other hand, after conditioning, the anode pattern consists of a large number of weak spots, all of about the same intensity, indicating a large number of emitters spread over the cathode surface.

Volt-ampere characteristics of the prebreakdown current are plotted according to the Fowler-Nordheim theory in order to yield typical surface parameters: critical field at breakdown, field-enhancement factor, emitting area, and cathode protrusion height. Analysis of the quantitative changes in these parameters before and after conditioning has led to further insight into the mechanism of gas-conditioning. The changes are demonstrated in the histograms of Figs. 5.3, 5.1, and 5.2 for critical field at the whisker tips at breakdown, the field-enhancement factor β, and the total emitting area A, respectively. The data included in these figures involves about 300 separate runs with electrode gap spacings ranging from 0.12 to 6 mm for the tungsten and from 0.2 to 2.0 mm for the copper. A run was accepted for inclusion only if the points were reproducible and its F-N plot was a straight line.

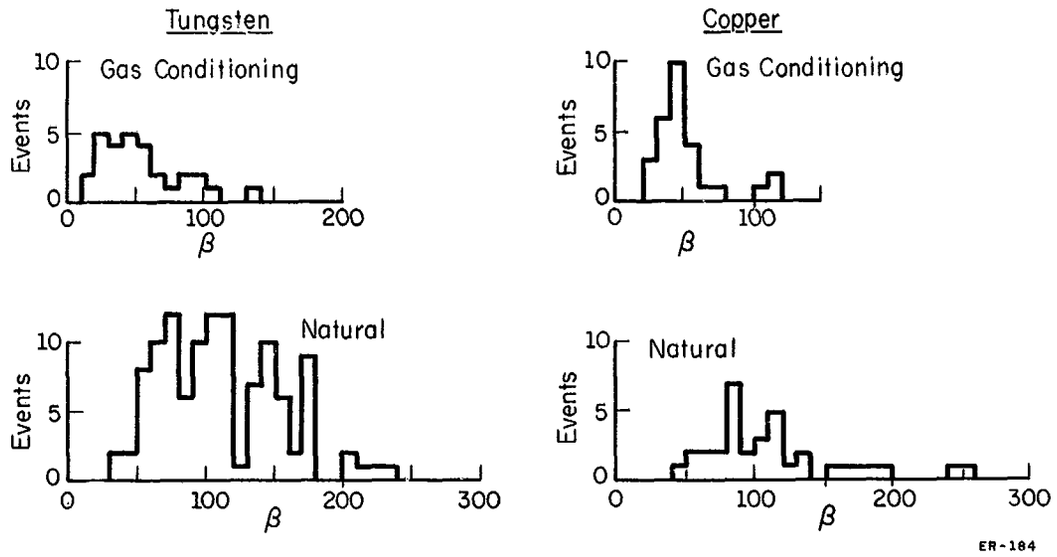


Fig. 5.1. Field enhancement factor.

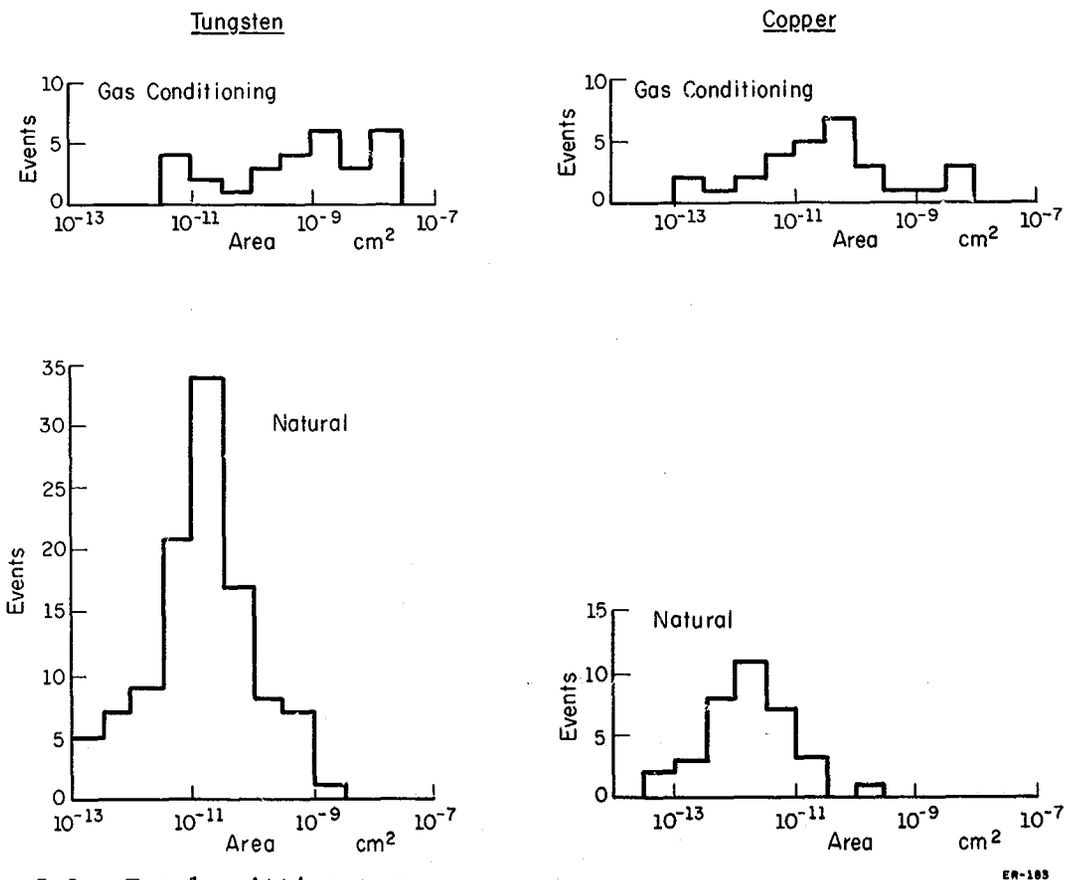


Fig. 5.2. Total emitting area.

(OVERLEAF BLANK)

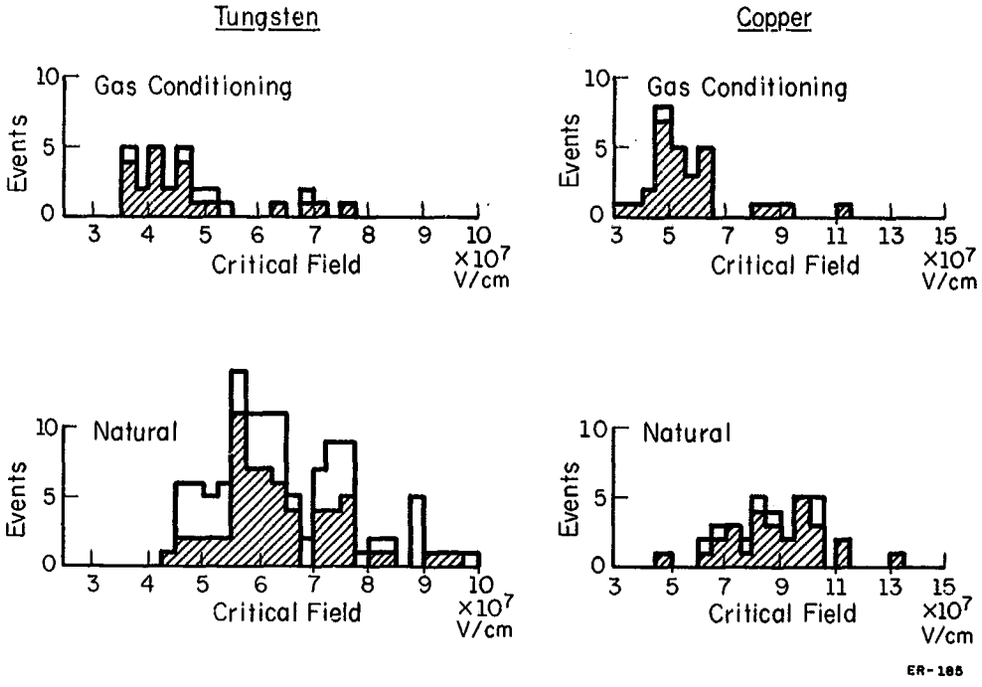


Fig. 5.3. Critical field at whisker tips. Shaded portions refer to runs terminated by a breakdown. Unshaded portions refer to a run in which the breakdown voltage point was assumed.

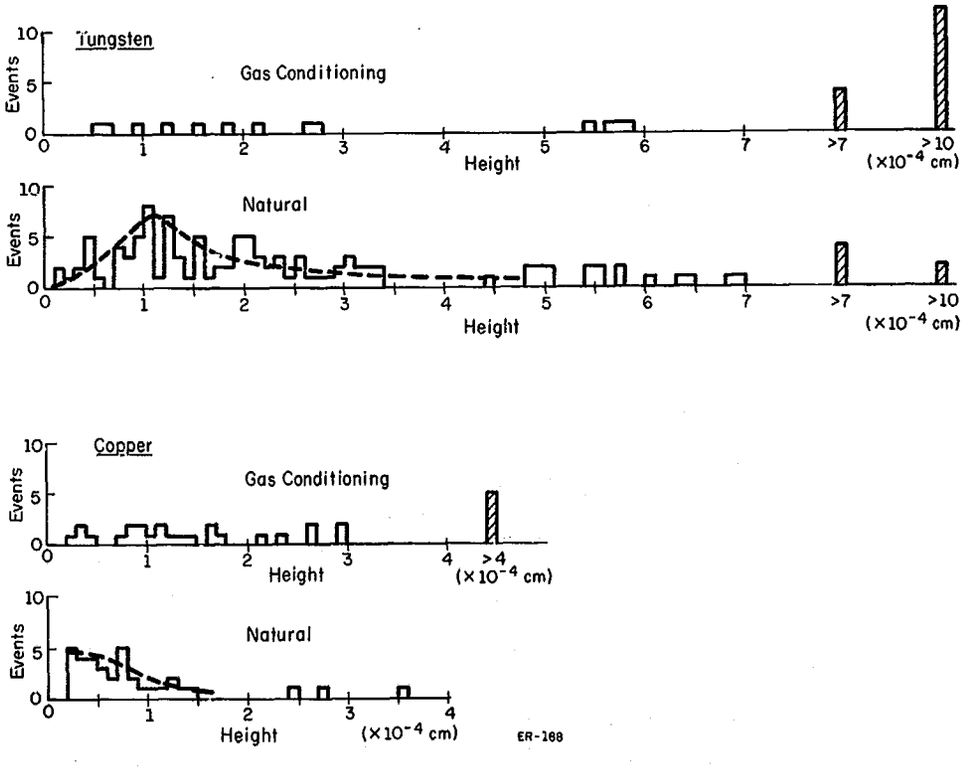


Fig. 5.4. Whisker height calculated from Eq.(1).

(OVERLEAF BLANK)

Figure 5.1, for the field enhancement factor, β, shows that for both copper and tungsten, gas conditioning reduces β by a factor of 2 to 3.

Figure 5.2, for total emitting area, as calculated from I/J, shows that gas conditioning increases the areas by one to two orders of magnitude. Despite this, the prebreakdown field-emission currents just below the breakdown-voltage points were found to be substantially lowered (by as much as an order of magnitude) by gas conditioning.

Figure 5.3, for the critical field at the whisker tips at breakdown, shows that the critical field for a given material is best represented as a distribution rather than a single discrete value. The average value depends on the material. Although gas conditioning reduces the average value of the field that can be sustained at the whisker tips, it was observed that the breakdown voltage for a fixed gap, was always increased by gas conditioning. The improvement in the voltage-holding capability was found for copper (at gaps smaller than 2 mm) to decline as the gap length was increased.

By utilizing Vibran's [4] formula relating field-enhancement factor β, whisker height h, and whisker tip radius r,

$$\beta = 2+h/r \approx h/r \approx h/(A/\pi)^{\frac{1}{2}} , \qquad (1)$$

the effective height h, of the single whisker which would give the V-I characteristics observed on a given run may be calculated. The histograms, Fig. 5.4,

for "natural" and "conditioned" whiskers calculated in this manner indicate that the effective heights of natural whiskers appear to follow a fairly compact distribution whose typical average value depends on the material. On the other hand, the effective heights of the conditioned whiskers are randomly distributed over a wide range. It must be noted that the above calculation of whisker height from the average β and the total emitting area, I/J, does not distinguish between single and multiple emitters. In a recent experiment, the number of major emitters on the surface of broad-area tungsten electrodes was determined by counting the transition-radiation bright spots (from 1 to 5) visible on the anode. The mean whisker height of a single whisker was then determined from the mean emitter area. This distribution is shown in Fig. 5.5 and indicates an average height for tungsten, of about 1 micron, in general agreement with determinations by several other methods.

The above observations may be interpreted as evidence of a qualitative difference between whiskers of different materials and between "natural" whiskers and "conditioned" whiskers of a given material.

5.1.2. Discussion of Results

In order to explain the qualitative and quantitative observations listed above, a short discussion of the gas-conditioning process seems appropriate. As reported by Alpert *et al.* [3] the decrease in β which results from gas conditioning may be explained on the basis of the blunting of a whisker tip due to ion bombardment. In addition to whisker blunting,

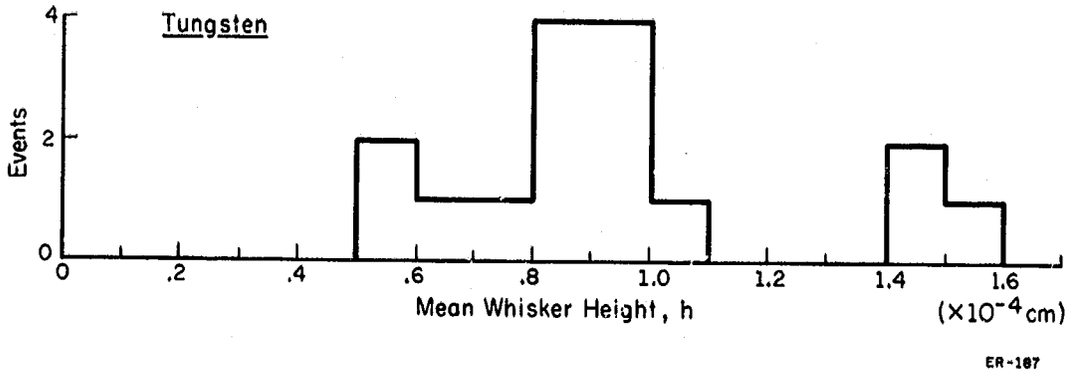


Fig. 5.5. Mean height of single tungsten whisker.

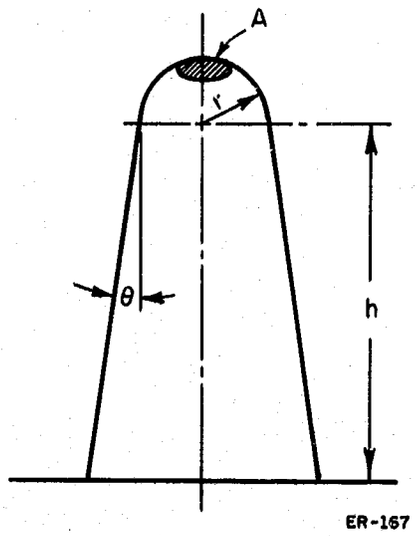


Fig. 5.6. Emitter Geometry.

(OVERLEAF BLANK)

sputtering of the cathode surface also occurs, with the consequent build up of new whiskers formed as stacks of sputtered particles. The formation of such whiskers due to ion bombardment on tungsten and due to electron bombardment on copper has been reported, for example, by Tomaschke et al. [5] and Maitland et al. [6], respectively. The nature of this process suggests that such whiskers will be abundant and will be randomly distributed over the entire electrode surface. Hence a large number of sites having low β contribute to the total emission current. The total emitting area, as calculated from I/J, should be expected to be considerably larger than the total emitting areas for natural whiskers. Because of the multiplicity of the conditioned whiskers, no real significance can be attributed to their heights as calculated by Eq.(1).

The observed fact that both the critical field and the prebreakdown emission current are lower for "conditioned" whiskers than for "natural" whiskers suggests that the former are probably less homogeneous than the latter. Consequently, the inferior heat conduction to the base and the increased resistance of the whisker would permit the melting temperature at the whisker tips to be reached at a significantly lower emission-current density, and a lower value of critical field than for natural whiskers. Even though the total emitting area for conditioned whiskers is greater than for natural whiskers, the total emission current just below the breakdown point is lower because of the reduced critical field. Breakdown might also be initiated at a weak point in the whisker

structure resulting from a mechanical failure or heating to the melting point at a restricted cross section, even though the temperature at the whisker tip is well below the melting point.

The formation of inhomogeneous whiskers on the cathode surface suggests an important limitation of the gas-conditioning process: The ultimate conditioning is obtained for a voltage where the rate of blunting is counterbalanced by the rate at which new sputtered whiskers are formed.

5.1.3. Theory relating critical field to whisker height of "natural" whiskers

The model [1] assumed here for a "natural" whisker is a truncated cone of half angle θ ($0 < \theta < 4^{o}$) and height h, terminated in a spherical tip of radius r as shown in Fig. 5.6. Cathode-induced breakdown is assumed to occur when the temperature at the tip reaches the melting point. The difference, ΔT, between the tip and the base temperature is

$$\Delta T = (\rho J^2 H^2/2K)+(JEH/K\phi^{\frac{1}{2}})6.9\times10^{-9}, \qquad (2)$$

where E is the field at the tip, J is the emission current density, ρ, K, and ϕ are the average values of the resistivity, heat conductivity and work function respectively, and H, a measure of the virtual protrusion height is given by

$$H = h(r/\theta)[h+r/\theta]^{-1} = h/(\beta\theta+1). \qquad (3)$$

In deriving Eq.(2) it was assumed that thermal balance obtains between heat generation due to Joule and Nottingham* mechanisms and heat loss to the base by conduction. In the two limiting cases, a perfect cylinder ($\theta = 0^{o}$), and a wide angle cone where $h \gg r/\theta$, Eq.(2) reduces to the forms quoted by Charbonnier et al. [7] The numerical factor in Eq.(2) was suggested by Chatterton [8] for an approximate calculation of the Nottingham heat contribution.

As long as the cone angle is small ($\theta < 4^{o}$) it is assumed that the height may be calculated from Eq.(1), i.e., the whisker belongs in the cylindrical-shape category.

Some interesting qualitative conclusions may be drawn from Eq.(2) on the assumption that cathode-induced breakdown occurs when the field E at the whisker tip reaches a critical value E_c, thus raising the tip temperature to the melting point: Eq.(2) predicts that for whiskers of a given cone half-angle θ, the critical field is not constant, but decreases with increasing whisker height, and a minimum value of the critical field is to be associated with a whisker of cylindrical shape and maximum height. Higher critical fields than this minimum are associated with shorter, or cone-shaped whiskers.

---

*The Nottingham heating effect should be taken into account only when the melting temperature of the material is lower than the Nottingham inversion temperature.

The dashed curve in Fig. 5.7a is the theoretical relationship between whisker height and critical field for cylindrical protrusions ($\theta = 0$), where average tabulated values of ρ, K, and ϕ for tungsten have been substituted into the equation. The solid-circle points are the values of critical field and height calculated from the experimental measurements on pure tungsten before any argon gas had ever been introduced into the system. The fact that these points generally lie to the right of the theoretical line lends support to the validity of our model for cathode-induced breakdown, and the reasonable hypothesis that geometrical differences in the whiskers demand a range of critical fields for breakdown.

The open-circle points are for runs after argon had been introduced. These runs were classified as "natural" because the measurements were preceded either by mechanically touching the electrodes together or by vacuum sparking. It is apparent that the argon conditioning had indeed modified the surface, reducing the critical field as explained above. The solid curve has been fitted at one point ($h = 0.95\ \mu m$ and $E_c = 6 \times 10^7$ V/cm) to correspond to the experimental observations.

The relationship between E_c and h demonstrated in Fig. 5.7a permits prediction of a minimum voltage, above which breakdown will occur for a given electrode pair. The prediction is based on nondestructive measurements of the V-I characteristics, taking into account the appropriate relationship between E_c and the whisker height. From the values of β and A, h is computed, and the corresponding E_c is found from Fig. 5.7a.

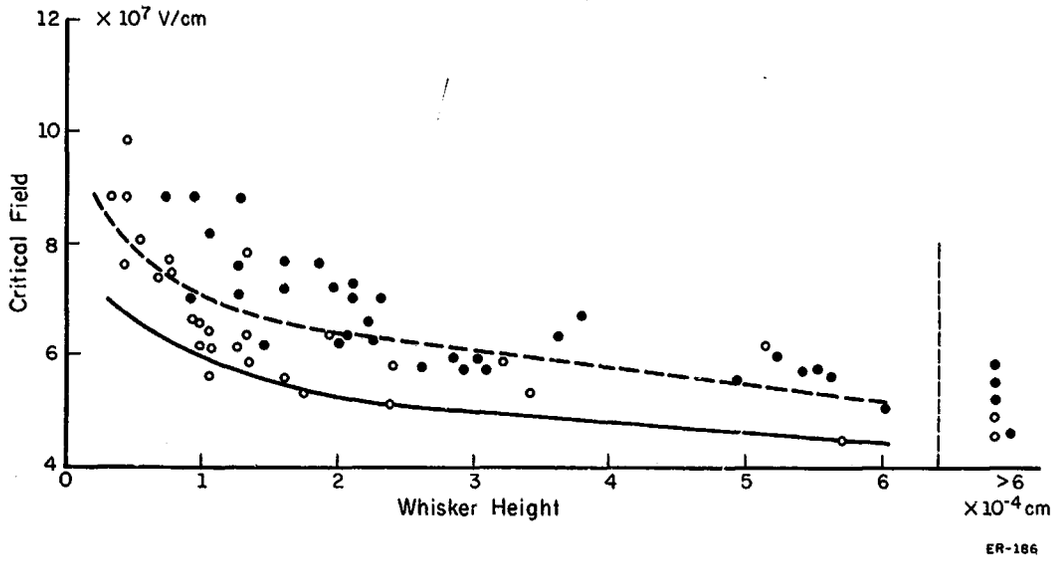


Fig. 5.7a. Height h of natural whiskers on tungsten calculated from total emitting areas. Curves are plots of Eq.(3). Dashed line assumes average tabular values for ρ, K and ϕ. Solid curve is fitted at one point to the experimental results. ● refer to runs made prior to introduction of argon; ○ for succeeding runs.

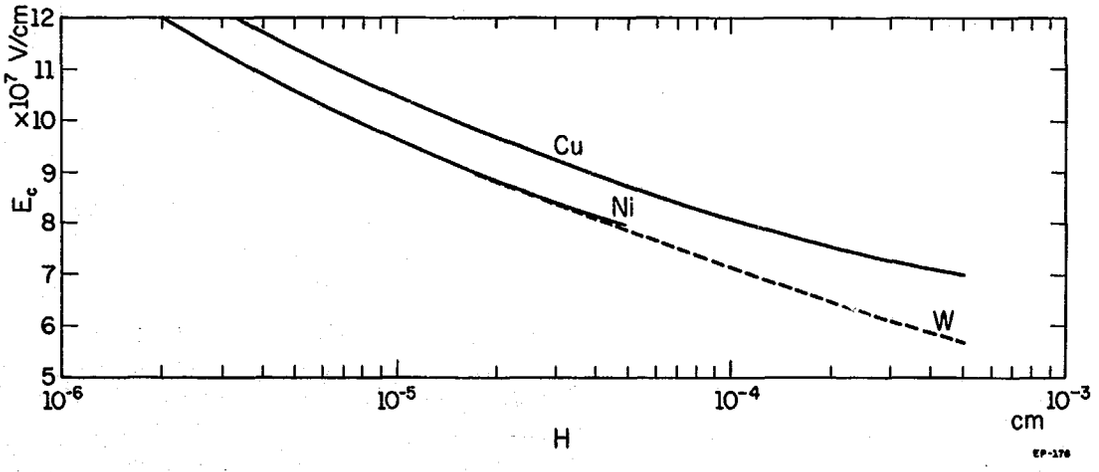


Fig. 5.7b. Virtual height H. Plot of Eq.(3) for virtual height using average tabular values for ρ, K and ϕ.

(OVERLEAF BLANK)

Then $V_{BD} = E_c(d/\beta)$, in which d is the electrode gap spacing. For a conical protrusion, V_{BD} would be higher than that computed above which refers to a cylindrical protrusion. If the protrusion cone-angle were known, the parameter H could be obtained from Eq.(3), and a better approximation could be obtained for the value of E_c from Fig. 5.7b. It might be noted that two orders-of-magnitude change in H results in only fifty-percent change in E_c. The observed extent of the range of E_c is illustrated in Table 5.1. The small change in E_c which is associated with a much greater variation in H indicates that although cathode breakdown may be induced by whiskers that vary considerably in shape and size the resultant effect on the critical field is limited to a narrow range.

Table 5.1 Range of values of E_c for natural whiskers.

	$E_{c\ min.}$	$E_{c\ max.}$
Cu	6×10^7 V/cm	13×10^7 V/cm
W	4.5×10^7 V/cm	9×10^7 V/cm
Ni*	9×10^7 V/cm	11.7×10^7 V/cm

*Values for E_c were obtained from Ref. 9.

5.1.4. Conclusions

In summary, for broad-area electrodes, we find that

a. Conditioning of W and Cu electrodes by argon gas increases the breakdown voltage. We observed a marked decrease in β, a small decrease

in E_c, and an increase in the emitting area. It also produces a change in the surface character.

b. The limitation of gas conditioning is due to the fact that the ultimate conditioning state is reached when the rate of whisker blunting is counter balanced by the rate of whisker formation as a result of sputtering.

c. The correlation between whisker heights and critical fields deduced from experiment is in substantial agreement with the theory of cathode-induced breakdown based on the melting of the tips of sharp protuberances on the cathode. This theory provides a criterion for breakdown which is based on nondestructive measurements of the pre-breakdown current and voltage.

E. M. Lyman
S. Y. Ettinger

5.2. Cylindrical Field-Emission Microscope

5.2.1. Observations

Preliminary observations of the conditions which facilitate whisker growth on single-crystal tungsten wire have been carried out utilizing a cylindrical field-emission microscope. As shown in Figs. 5.8 and 5.9, this consists of a long tungsten wire, 0.067mm in diameter, as a cathode; a conductive fluorescent glass screen as an anode; a 25-mm-long, stainless-steel tube, 2.79 mm in diameter, which can be moved along the

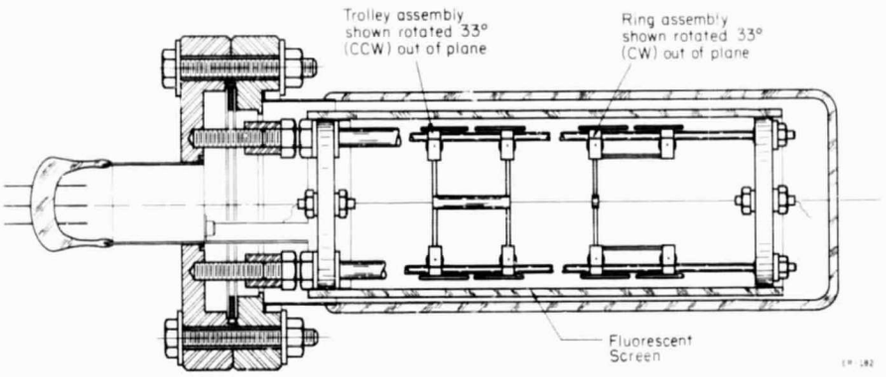


Fig. 5.8. Cylindrical field-emission microscope.

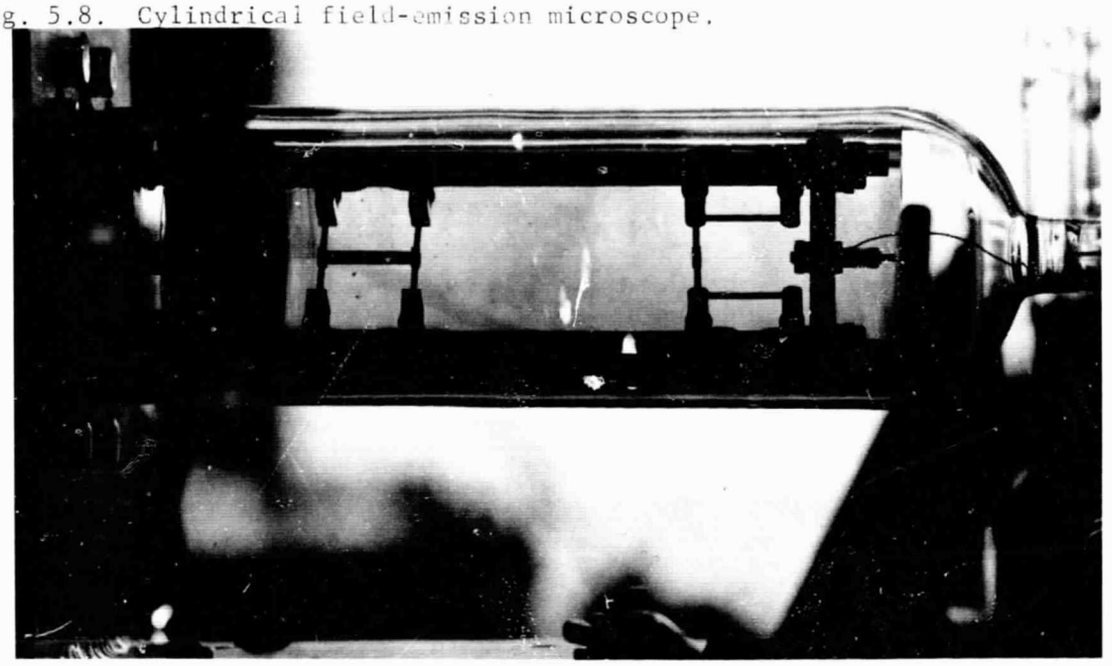


Fig. 5.9. Field-emission patterns on the screen of the cylindrical FEM.

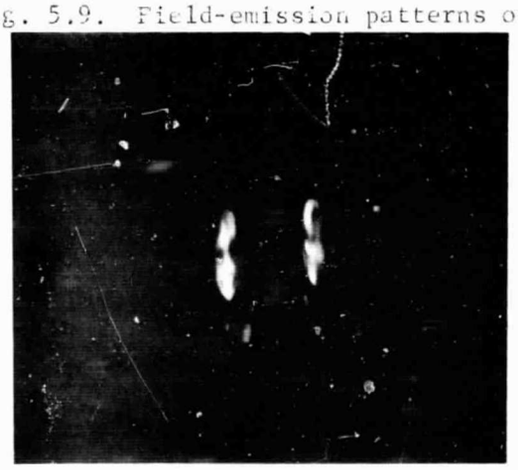


Fig. 5.10. Pattern of a crystalline tungsten whisker.

Fig. 5.11. Fowler-Nordheim curve autometically plotted by BETA 3.

(OVERLEAF BLANK)

wire by means of an external magnet serving as an optional anode; and a 3mm-long stainless-steel ring, 2.79 mm in diameter, which is also movable along the wire, used for measuring field-emission currents from single emission sites.

The objectives of the present work are:

a. To study the contribution of a metallic anode to whisker formation.

b. To compare the rate of whisker formation under identical conditions on surfaces exposed to the metallic anode on one hand and to the fluorescent glass screen on the other.

c. To study the combined effect of temperature and electric field on whisker growth. Here again, the field can be applied either between the wire and the metal tube or between the wire and the screen.

When the wire is heated its temperature is determined from resistance change measurements.

The following conclusions are derived from preliminary results obtained thus far:

a. When the cathode is at room temperature, the formation of whiskers within a period of 1 hour necessitates the presence of the metallic tube as an anode and a threshold emission current of the order of 10^{-6} A. Application of the same cathode electric fields (of 1.4×10^{6} V/cm) when the screen is used as anode for a 1 hour

period did not result in the formation of new protrusions even though the total emission current in this case was 10^{-4} A.

b. Heating the cathode to 750°C established favorable conditions for whisker formation when a field of 4.7×10^{5} V/cm was applied between the wire and the screen. However, in this case too, a threshold emission current of approximately 1×10^{-6} A was necessary to cause whisker growth during the 1 hour observation period.

c. The protrusions which are formed in either of the processes described above, are sometimes of a genuine metallic nature, in which case an image of the crystalographic structure of the emitting tip can be observed on the screen, as demonstrated in Fig. 5.10. On other occasions, however, the new whisker is formed due to adsorption or reorientation of contaminants on the wire surface. The emission pattern in such cases does not display a characteristic fine structure, and by heating the wire, one may cause the emission site to disappear *via* desorption.

5.2.2. Discussion

The derivation of typical parameters (e.g., β, etc.) for whiskers produced by contaminants yields ambiguous results because of the uncertainty in the exact value of the work function. In principle, these parameters can be determined when the emission current is measured at two

different temperatures. [10] Noting that the temperature dependence of a field emission current is given by [11]

$$J_T = J_0 \pi P/(\sin \pi P), \tag{1}$$

in which

$$P = kTt(y)(\phi^{\frac{1}{2}}/\beta E_w)/9.76\times10^{-9}, \tag{2}$$

where J_T and J_0 are the current density at temperatures T and 0^oK respectively; t(y) is a tabulated function of y [10]; ϕ is the work function in eV; β is the field-enhancement coefficient; E_w is the average electric field at the wire surface.

At room temperature $J \approx J_0$. Assuming constant values for the field, work function and emitting area, the following relationship between the field and work function is obtained, viz,

$$(i_T - i_0)/i_0 = \Delta i/i_0 = \pi P(\sin \pi P)^{-1} - 1 \approx (1/6)(\pi P)^2. \tag{3}$$

From the measurements of the slope in the Fowler-Nordheim plot a second relation between the field and work function is available, namely:

$$|m| = Bs(y)\phi^{3/2}/\alpha\beta, \tag{4}$$

in which $\alpha = E_w/V$, with V being the applied voltage, m the slope; B a constant, and s(y) another tabulated function of y [11]. A simultaneous solution of Eq.(3) and (4) gives the values of β and ϕ for the site under

consideration. From these, the area and height of the protrusion can be derived [12]. The calculated change of the emission current over the temperature range for which Eq.(3) is valid, is less than 10%. To make a meaningful measurement requires extremely stable emission conditions where the current fluctuations are less than 1%. In practice, however, this stringent requirement cannot be met since the measured current fluctuated by approximately 20%. Hence, the parameters of a contaminated site could not be determined by this method.

In those cases where the work function is known, namely uncontaminated tungsten whiskers, a rough estimate of the whisker shape can be deduced. This is achieved by comparing the results for the whisker height as calculated by two independent methods - the image size on the screen and the Fowler-Nordheim plot. For a cylindrical projection tube Brodie [13] has shown that the whisker height h can be derived from the length, $2r_o$ of minor axis of the elliptical image, namely

$$h = R\{\exp[(r_o^2/R_2^2)\log(R_2/R_1)]-1\} \approx R(r_o^2/R_2^2)\log(R_2/R_1) \qquad (5)$$

In this equation, R_1 and R_2 are the radii of the center wire and the anode screen, respectively. In the present experiment $R_1 = 3.3\times10^{-3}$cm and $R_2 = 2.98$ cm, so that Eq.(5) becomes

$$h \approx 1.1\times10^{-3} r_o^2 \text{cm}. \qquad (6)$$

Knowing the values of β and the emitting area A, as obtained from the FN plot, and using calculated results for the dependence of β on the whisker dimensions for various basic geometries [8,13] (e.g., spheroid, truncated cone terminated with a spherical tip, cylinder with a spherical tip etc.), corresponding whisker heights for each geometry can be deduced.

In one specific case the data were: $r_o = 0.2$ cm, $\beta = 46$; $A = 3.4 \times 10^{-13} cm^2$. The value of the whisker height obtained from the image size is

$$h = 1.1 \times 10^{-3} \times 4 \times 10^{-2} = 4.4 \times 10^{-5} cm. \quad (6)$$

The heights calculated for the same site assuming different geometries are summarized in Table 5.2. Comparing these values with the value obtained from the image calculations indicates that in this case the protrusion is best described as having a conical shape.

Table 5.2

Protrusion Geometry	Protrusion height from β and A	Protrusion height calculated from image size
cone	3.9×10^{-5} cm	4.4×10^{-5} cm
cylinder	1.45×10^{-5} cm	
spheroid	9.6×10^{-5} cm	

S. Y. Ettinger

5.3. Computer Programs for Calculating Surface Parameters and Drawing Fowler-Nordheim Plots

5.3.1. BETA 3

One of the reasons why it has been feasible to make the statistical studies reported in section 5.1 has been the development of a computer program which not only computes the desired parameters but also automatically plots out the Fowler Nordheim curves, as shown in Fig. 5.11.

Determination of electrode surface conditions during voltage breakdown in high-vacuum systems can best be accomplished through studies of emission current characteristics over a range of voltages. According to the Fowler-Nordheim theory, if emission current I and voltage V are graphed in the form $X=1/V$, $Y=\log_{10}(I/V^2)$, the resulting plot will be approximately linear. From the slope and y intercept of such a plot, several important parameters may be calculated:

a. β (the field-enhancement factor)

b. N (a factor to convert potential difference into field intensity)

c. critical field E_c

d. effective emitting area A

e. height h and radius of curvature r of tips of emitting whiskers.

BETA 3 is a computer program which provides a convenient method for high-speed calculation of these factors from experimental current-voltage data.

The experimental current-voltage data are read in by BETA 3 and graphed as described above. A line is fitted to the data using a least squares criterion on the y parameter so that the slope and y intercept of the line are found. From the slope, N may be calculated according to the relation [11]

$$m = B \phi^{3/2} s(y)/N$$

where m is the slope, ϕ is the work function of the electrode metal, B is a constant, and s(y) is a function of the parameter y which is in turn a function of N and V. The function s(y), however, is a complicated one involving elliptic integrals and is difficult to solve on the computer. BETA 3, therefore, uses a quadratic approximation for s(y) that agrees with the values tabulated in Good and Müller [11] within +0.6% to 0.04%. The factor β is calculated according to the relation

$$\beta = Nd,$$

where d (normally the gap spacing between parallel electrodes) is a constant for any given set of data such that

$$V = F_{av} d,$$

where F_{av} is the average field intensity at the cathode surface. The critical field and the average field are calculated in the standard manner

from N and the breakdown voltage. The effective emitting area is calculated from the slope and the y intercept according to the equations developed by Charbonnier and Martin.[14] From the area and β, the height and radius of curvature of whiskers are calculated as described in the CSL Progress Report of April 26, 1968.

At low currents (10^{-11} to 10^{-9} A), there may be a systematic error in the experimental current-voltage data due to the stray resistance in parallel with the electrodes. The program tests for this effect and eliminates, prior to the calculation of the line, any points which are noticably affected by this residual current.

A section of the program was developed to allow on-line changes in various parameters used in the calculation of N and β. This section enables changes to be made by the user in ϕ and d to allow for the natural uncertainty in these parameters. Also, typographical errors in the data tape may be corrected at the time of processing. This section may be used to subtract the residual current due to the stray resistance of the system or to eliminate data points too divergent to be handled by the main section of the program.

A final section of the program was developed to allow storage and retrieval of data in binary format on magnetic tape for future reference.

5.3.2. BETA 4

Another program, BETA 4, is also under development. BETA 4 is being developed in the Illar system on the CSL computer, whereas BETA 3

was developed in the Fortran '60 system. The Illar system will allow several improvements in the Program:

a. more efficient and flexible use of data stored on a magnetic file tape.

b. more efficient use of computer time.

c. higher reliability in processing of data.

d. a greater degree of error testing.

e. fuller automatization of processing.

The modular nature of the programs in the Illar system will also simplify modifications in processing technique as the mathematical theory is modified and refined.

Programs have been developed to store and retrieve data in binary format on magnetic tape, to read data and test for typographical errors, to bring out the results of computation, and to communicate with the program user via the video display unit and the light pen. Work is now under way on the processing sections of the program which are basically similar to those used in BETA 3.

5.3.3. PACKPEN

Although most of the programs developed are applicable only to the high-voltage breakdown experiments, the program for communicating via the light pen may be useful to other groups. This program, PACKPEN, allows the user to display on the video display unit a heading and up to 60 statements. The heading may be up to 60 characters long and the statements may

be as long as 16 characters each. Associated with each statement is a light button. By pressing the light pen to various light buttons, the user may transmit information to his program. PACKPEN is a Fortran program, but may be easily converted to assembly language. PACKPEN will be used in BETA 4 to select program options, to select data sets to be recalled from the file tape, and to initialize various processing parameters.

J. Hansen

5.4. References

1. E. M. Lyman, D. A. Lee, H. E. Tomaschke and D. Alpert, Proc. II Int. Symp. Insul. High Volt. Vac., 33 (Sept. 1966).
2. D. Alpert, D. A. Lee, E. M. Lyman and H. E. Tomaschke, J. Vac. Sci. Tech. 1, 35, (1964).
3. D. Alpert, D. Lee, E. M. Lyman and H. E. Tomaschke, J. Appl. Phys. 38, 880, (1967).
4. G. Vibrans, J. Appl. Phys. 35, 2855, (1964).
5. H. E. Tomaschke and D. Alpert, J. Vac. Sci. Tech. 4, 192, (1967).
6. A. Maitland, Brit. J. Appl. Phys. 16, 1951 (1965).
7. F. M. Charbonnier, C. J. Bennette and L. W. Swanson, J. Appl. Phys. 38, 627, (1967).
8. P. A. Chatterton, Proc. Phys. Soc. 88, 231, (1966).
9. I. Brodie, J. Vac. Sci. Tech. 3, 222, (1966).
10. R. D. Young and E. W. Müller, J. Appl. Phys. 33, 91 (1962).
11. R. H. Good and H. E. Müller, Handuch der Physick, Vol. 21, p. 185, Springer-Verlag, Berling, 1956.
12. E. M. Lyman et al., CSL Progress Report, Sept. 1 967-Feb. 1968, University of Illinois, p. 79.
13. I. Brodie, J. Appl. Phys. 35, (1964), 2324.
14. F. M. Charbonnier and E. E. Martin, J. Appl. Phys. 33, 1897 (1962).

W. J. Bouknight
D. H. Cooper
J. D. Gooch
G. R. Karr
B. D. Kirkwood
H. W. Knoebel
H. V. Krone
J. L. Myers, Jr.
D. O. Skaperdas

6.1. Relativity Satellite

6.1.1. Data Reduction

A method of data reduction that utilizes spin-axis precession rates that are periodic with the orbit nodal-regression frequency has been shown to be capable of accurately measuring the general-relativity effect in the precession of a gyroscope satellite in an inclined orbit. [1] The spin-axis precession of a six-mirrored satellite having nominal physical parameters and optically tracked by a network of 12 tracking stations was simulated by digital-computer programs. The results of these computer simulations indicated that, for the given satellite design and available tracking stations, the percentage error of the relativity measurement would be minimized for a circular orbit at 583 km altitude, 31° inclination and satellite spin axis initially directed approximately toward the vernal equinox. Moreover, deviations from these optimal values would not seriously degrade the measurement accuracy; for example, for a desired relativity measurement of 2% accuracy, an initial spin axis misalignment of $\pm 3^{\circ}$ may be tolerated.

The techniques of spin-axis readout and data analysis described in Ref. 1 do not require the absolute minimization of precessions due to the extraneous torques of gravity gradient, solar pressure, and aerodynamic

†This work was supported by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

drag, but only require that these precessions be reduced to a level such that stationarity of the spin axis may be assumed for one-day intervals. The measurement of each of these extraneous effects is one of the by-products of the data analysis.

By running the relativity satellite experiment for periods longer than one year, the measurement accuracy may be continually improved until the effects that were neglected in this analysis of a one-year experiment become significant.

J. L. Myers
H. W. Knoebel

6.1.2. Completion of Study

The favorable results of the above simulations extend and strengthen the conclusions of Ref. 2, the final report on the feasibility study of a relativity experiment started under NASA Grant NSG 503. Though one or two minor investigations are continuing, the feasibility study may be regarded as having been most successfully completed. A proposal is now being prepared to solicit support for the implementation of payload fabrication, launch, and ground-based supporting activities to conduct the measurement of the relativistic precession.

D. H. Cooper
H. W. Knoebel

6.2. Gas-Surface Interaction Satellite

In the previous CSL Progress Report [3] a proposed gas-surface interaction satellite experiment was described which would allow the determination of the absolute atmospheric gas density and a two-parameter gas-surface interaction model. These three quantities would be obtained from the measurements of drag force, spin-rate slow down and precession rate of a passive, spin-stabilized satellite. The National Aeronautics and Space Administration at the George Marshall Space Flight Center has taken interest in our work and is now providing support for further study. Following is a brief outline of the work we propose to perform in the following year.

6.2.1. Study of Satellite Aerodynamics and the Gas-Surface Interaction

A. A review of gas-surface studies both experimental and analytical is of course necessary to put our proposed work in perspective.

B. We must develop an analytic expression for the drag, slow down rate and precession of our proposed satellite incorporating appropriate gas-surface interaction models. We will first develop this expression using the hypothermal approximation until our work warrents more refined expressions.

C. An optimum shape for the proposed satellite is thought to exist such that maximum accuracy is obtained in the measurement

of the three parameters. The possible satellite shape has many constraints which we must take into consideration in the study.

D. Besides the primary objective of determining the gas density and two gas-surface interaction parameters we wish to study the dynamical effects of degassing and the meteorite degradation of the surface. We shall also weigh the advantages of measuring the gas density of various altitudes. Finally, the possibility of making a measurement of a general relativity effect [1,2] shall be investigated.

6.2.2. Measurement Methods and Accuracy

The Aerospace Group at CSL has recently completed an extensive study of the accuracy and feasibility of a passive technique for measurement of the precession of a satellite (see ref. 2). A similar effort will be undertaken with respect to the drag, spin rate slow down and precession of the proposed aerodynamic satellite.

6.2.3. The Satellite System

The study of ref. 2 includes an optimization of the satellite system parameters (i.e., orbit inclination, eccentricity, satellite size, etc.). Such a study will be paralleled in the proposed work on the aerodynamic satellite. Included in this study will be an analysis of spin-up and launching techniques.

6.2.4. Future of the Aerodynamic Satellite Work

Within the period of the next year we plan to illustrate the feasibility of the proposed satellite experiment. At that time the degree of participation of CSL in the launching of the satellite and collection of data shall come under consideration.

6.2.5. Literature Study

Work has been initiated on this study by a review of current experimental and analytical work on gas-surface interactions (see section 6.2.1 part A). Considerable information on this topic was gathered by Professor S. M. Yen and G. R. Karr at the Sixth International Rarefied Gas Dynamics Symposium held at the MIT Campus in Cambridge, Massachusetts, July 22-26, 1968. They found that there are no accurate laboratory gas-surface interaction studies being performed in the 1-to-10-eV energy range corresponding to earth-satellite velocities. This is the energy range in which CSL plans to obtain accurate data which would be of immediate engineering value. Considerable discussion of gas-surface interaction models was presented at the Symposium. We plan to analyze the various models and determine which to use in the study of the dynamics of the proposed aerodynamic satellite.

G. R. Karr

6.3. Infrared Fourier Interferometry

A far-infrared interferometer was recently acquired from the

Research and Industrial Instruments Company, one of the purposes being to investigate possible transmission windows in the 500 to $48 \times 10^3 m^{-1}$ range. The main advantage of interferometers over spectrometers in this range is associated with flux-gathering ability, resolution, and noise-limited detectors. An interferometer simultaneously operates on all of the m spectral resolution elements within a large spectral range during a given data-processing time, whereas the spectrometer serially operates on each of these m spectral resolution elements within this range. This advantage, known as Fellget's advantage, or the multiplex principle, results in a gain in signal-to-noise ratio of a factor of $\sqrt{m}$ when the detector is noise limited, when both instruments have the same flux-gathering ability, and operate for the same length of time. Additionally, from a practical standpoint, for instruments of the same flux-gathering ability and resolution, spectrometers employing narrow slits must be much larger than interferometers employing circular apertures.

A disadvantage of interferometers is that their output consists of an interferogram or autocorrelation function, which is the inverse Fourier transform of the power spectrum; hence, an analog or digital computer must be used to obtain the power spectrum. The recent development of fast-Fourier-transform programs (Tukey-Cooley program, [4,5] for example) considerably mitigates this disadvantage.

The laboratory instrument is a Michelson-type interferometer, of modular construction, employing reflecting optics, and employing a stepping

motor for moving one of the mirrors in steps of adjustable distance and time duration. The energy source consists of a 125-W, high-pressure, water-cooled, mercury lamp and the present detector is a diamond-windowed Golay cell (specific detectivity approximately 1.7×10^7 $mHz^{\frac{1}{2}}/W$). Future plans are to increase the specific detectivity by an order of magnitude with the use of a carbon bolometer detector. The source is chopped by a synchronous motor driven at 12.5 Hz and the signal from the Golay detector is synchronously detected, integrated and sampled by an analog-to-digital converter, a sample being taken each time the moveable mirror is stepped.

The total mirror travel is 50 mm, giving a total path difference of 100 mm so that for two-sided interferograms the spectral resolution is 2 mm of travel for 20 m^{-1} and for one-sided interferograms the resolution approaches 10 m^{-1}. Due account must be taken of spectral resolution and interferogram sampling; RC filtering in the synchronous detector and sampling duration; spectral resolution and mirror travel; and resolution and solid angle subtended at the interferometer by the source and detector.

Three initial experiments of varying spectral resolution of water vapor in air were performed. The total travel of the moveable mirror was successively increased in each experiment from 2 to 4 to 6 mm, giving expected spectral resolutions of 500 m^{-1}, 250 m^{-1} and 130 m^{-1} respectively. Double-sided interferograms were taken; i.e., the moveable mirror was translated an equal distance from the equal-path-length position so that the main peak of the interferogram occurred at the exact center. Figure 6.1

shows one such interferogram. The efficiency of the beam splitter, which itself acts like a Fabry-Perot interferometer, is high enough for use in the range 4×10^3 to 48×10^3 m^{-1}.

The corresponding power spectra, obtained by performing fast Fourier transforms of the above interferograms on the CDC 1604 computer, are shown in Figs. 6.2, 6.3 and 6.4. Corresponding portions of these spectra are shown enlarged in Figs. 6.5, 6.6 and 6.7, in which the increasing spectral resolution is clearly seen.

D. Skaperdas

6.4. References

1. J. L. Myers, "Optimization of a Gyroscopic Satellite General-Relativity Experiment," Coordinated Science Laboratory, Report R-396, September, 1968.

2. D. H. Cooper, G. R. Karr, J. L. Myers, and D. Skaperdas, "A Proposed Test of the Einstein Theory of Gravitation by Means of an Unshielded Orbiting Gyro Using Passive Telemetry," Coordinated Science Laboratory Report R-378, May, 1968.

3. Progress Report for September, 1967 through February, 1968, Coordinated Science Laboratory, University of Illinois, pp. 107-116.

4. W. T. Cochran et al., "What is the Fast Fourier Transform?" Proc. IEEE 55, 1664.

5. J. W. Cooley et al., "Historical Notes on the Fast Fourier Transform," Proc. IEEE 55, 1675.

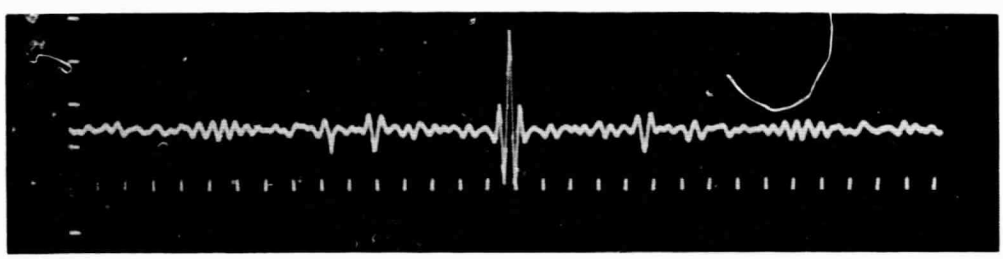


Fig. 6.1. Interferogram of water vapor.

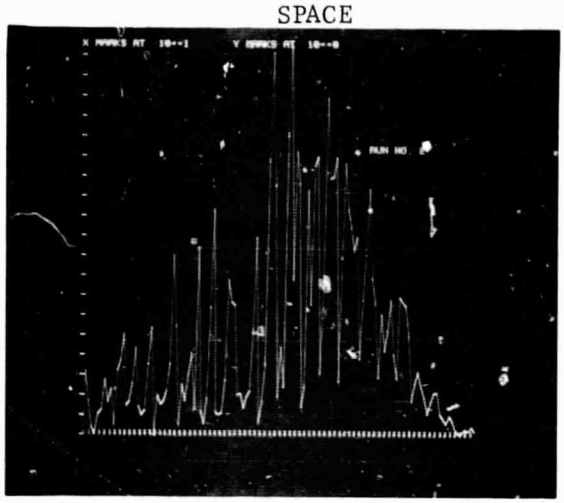


Fig. 6.2.

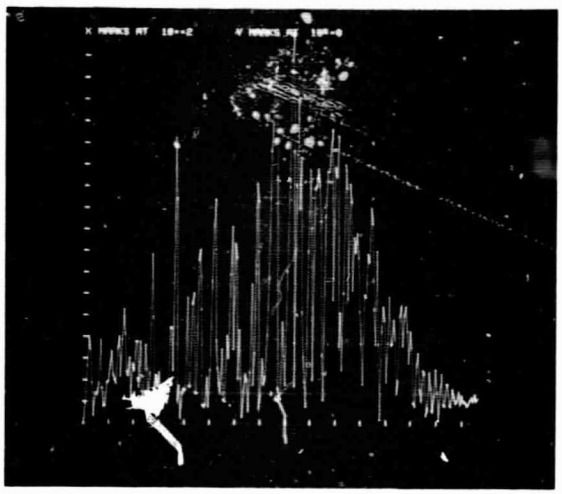


Fig. 6.3.

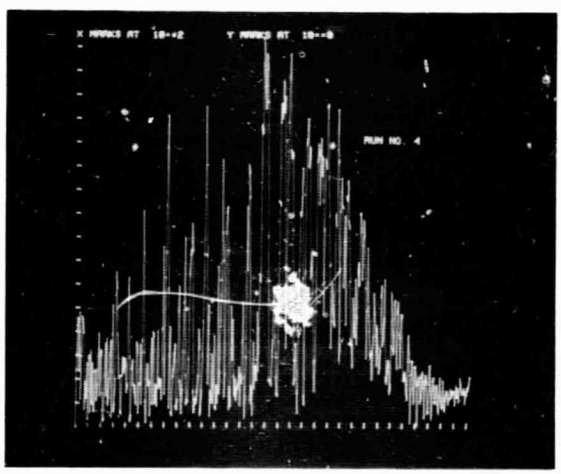


Fig. 6.4. Power spectra of increasing resolution for water vapor.

(OVERLEAF BLANK)

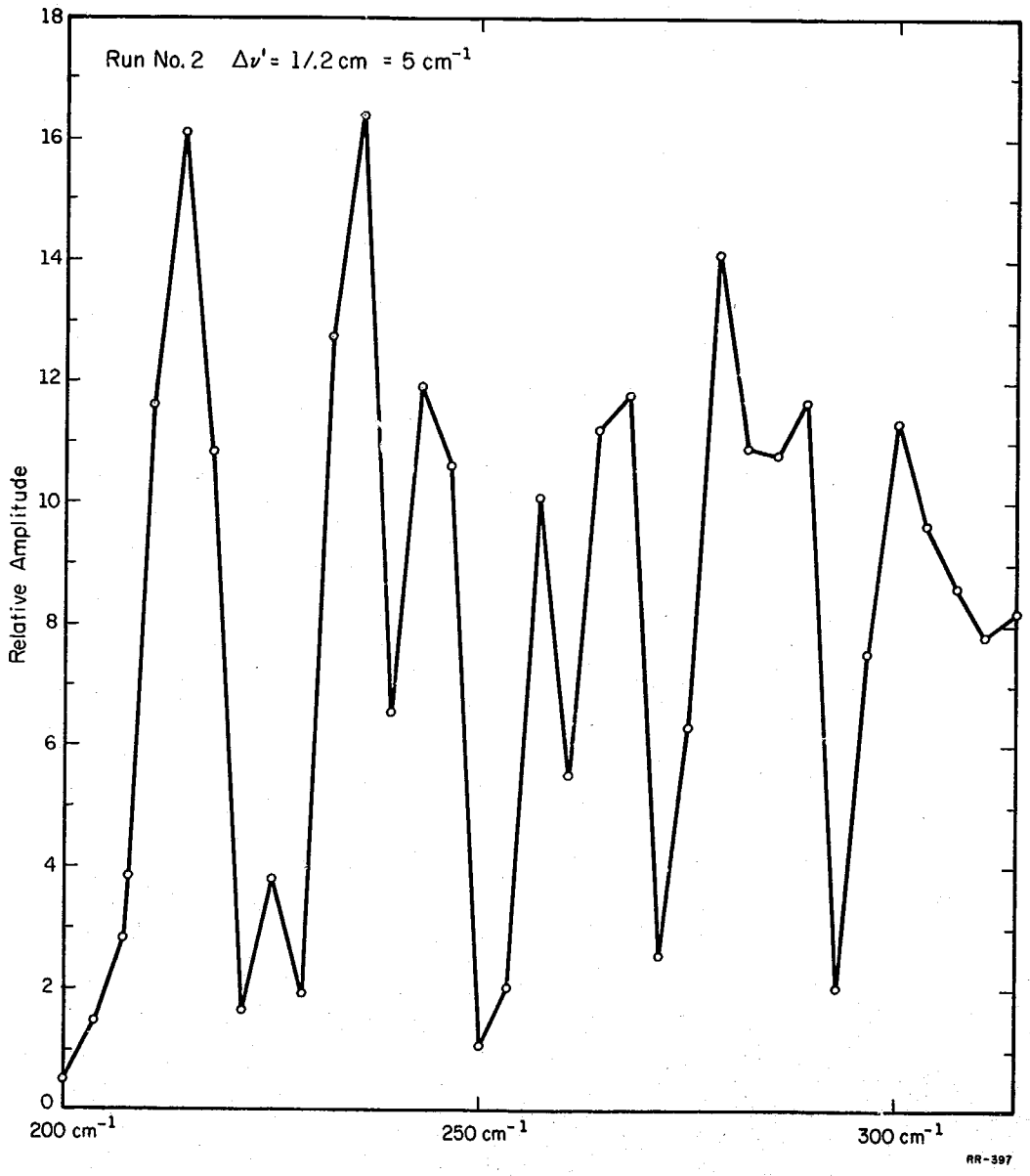


Fig. 6.5. Expanded power spectrum of water vapor at resolution of $500m^{-1}$.

(OVERLEAF BLANK)

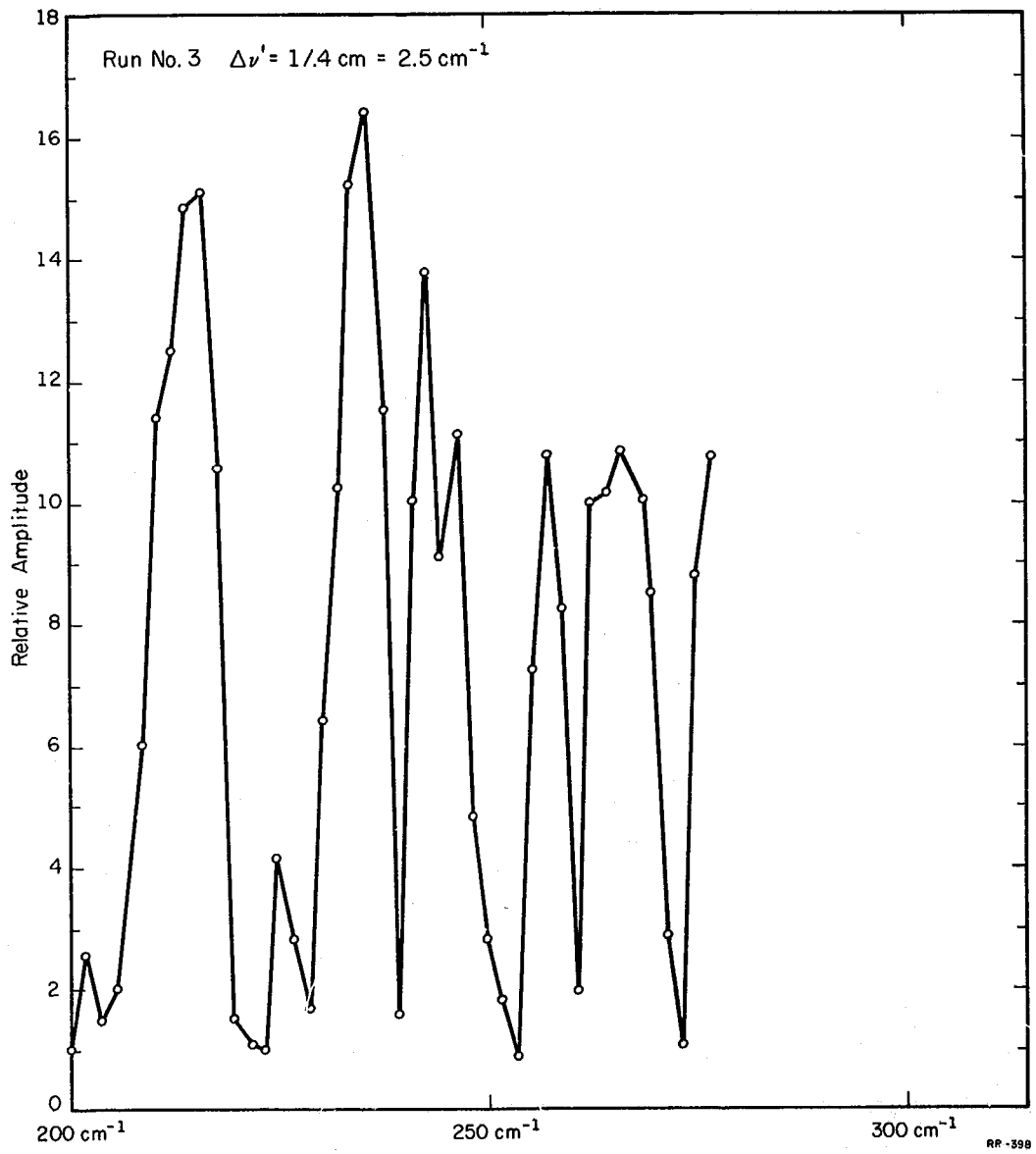


Fig. 6.6. Expanded power spectrum of water vapor at a resolution of 250m^{-1}.

(OVERLEAF BLANK)

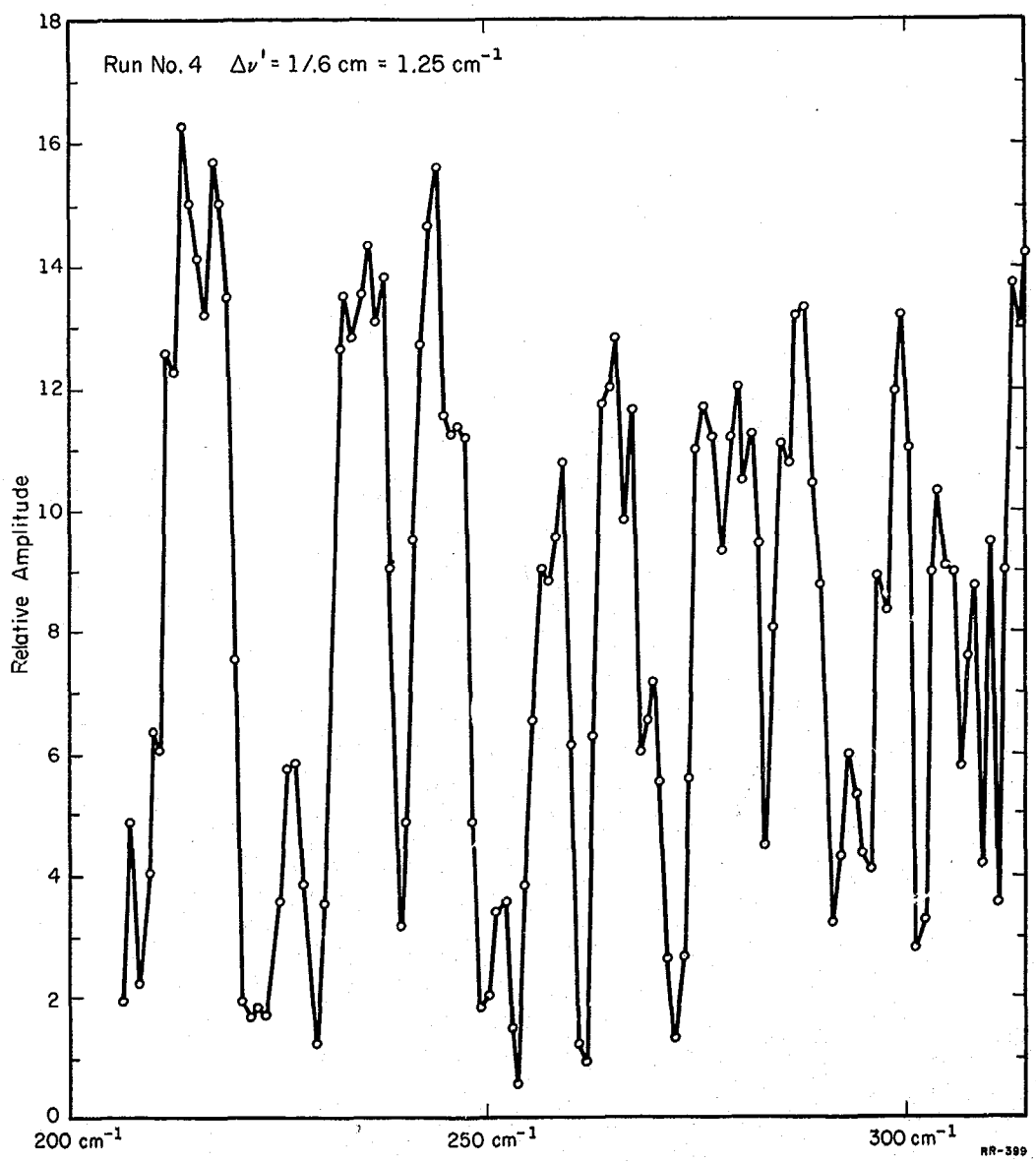


Fig. 6.7. Expanded power spectrum of water vapor at a resolution of 130m^{-1}.

(OVERLEAF BLANK)

7. SEMICONDUCTOR PHYSICS[†]

W. D. Compton
R. N. Peacock
H. V. Krone
J. P. Desroches
D. Skaperdas
F. Steinrisser
C. E. Jones
L. Schein
D. E. Cullen
E. S. Johnson
J. T. Jacobs

7.1. Introduction

Semiconductor devices normally require ohmic contacts for successful operation. This is often accomplished by doping the semiconductor, in the region of the contact, to a high level and attaching the metal contact to this doped region. A metal-degenerate semiconductor contact is thereby formed. The present research on such contacts is intended to provide information of a fundamental nature on the nature of the conductance process and the electron-hole coupling in these junctions.

Defects, either impurities or the fundamental lattice defects, often determine the properties of semiconductors. Irradiation with high-energy particles is a convenient way of introducing defects into solids. The recombination luminescence has been shown to be a sensitive indicator of the nature of the recombination center. Studies of the luminescence are underway, in an effort to determine the nature of a variety of the defects that are introduced into the material by irradiation.

The temperature dependence of the oscillatory effect in p-i-n junctions doped with Au or Co has been investigated. These structures have great possibilities as infrared detectors.

[†]Supported by the National Aeronautics and Space Administration under Grant NsG 228-62 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

The following sections present the results of studies in each of these areas.

7.2. Metal-Semiconductor Tunneling

7.2.1. Pb, In--Germanium Junctions

Preliminary results on the tunneling characteristics of germanium-metal contacts, with lead or indium used as the metal were reported in the previous Progress Report [1] and in a letter that has been submitted for publication to Physical Review. [2] Subsequent work has allowed a quantitative comparison between the one-electron-model predictions of conductance vs bias [3] and results for Sb-doped Ge. The height of the tunneling barrier was found to be 0.63 ± 0.03 eV from capacitance vs bias measurements at 77°K and was independent of contact metal (In or Pb), doping level, and dopant (Sb or As). The doping level was determined from resistivity measurements. [4] Figure 7.1 shows an absolute comparison between theory and experiment. Note that the effective mass is the only adjustable parameter.

Phonon-assisted tunneling was observed in Sb-doped Ge with approximately the same strength as in As-doped Ge. Figure 7.2 shows conductance dI/dV and d^2I/dV^2 vs bias for such a junction. With In as contact metal, the superconducting energy gap of In at $T=2^{\circ}$K and the increase in conductance at bias values corresponding to the energies of the four zone-boundary phonons can be easily seen. The LA-phonon is the strongest, in agreement with selection rules.

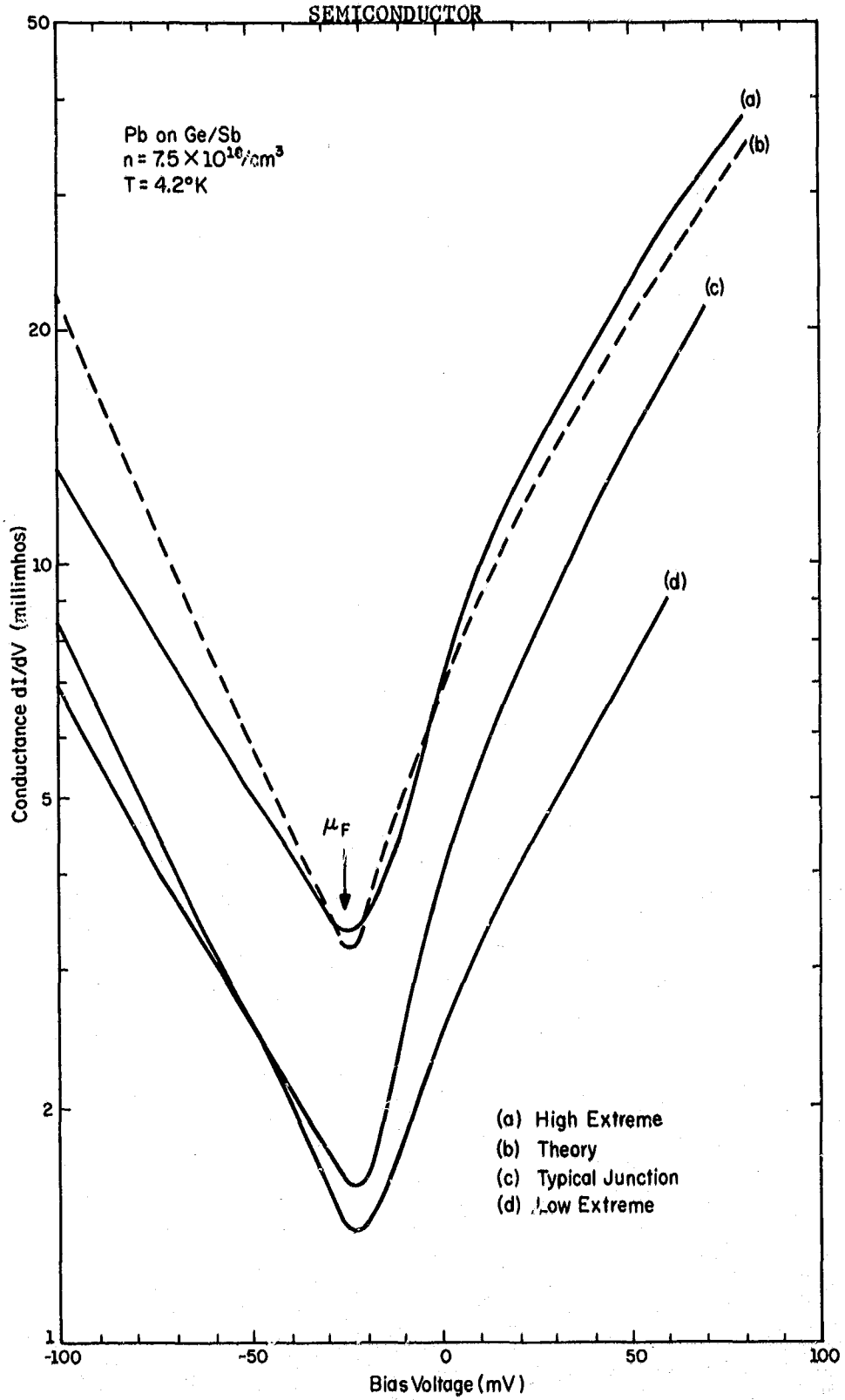


Fig. 7.1. Comparison between 3 experimentally-measured conductance curves on $n=7.5\times10^{18}/cm^3$ Sb-doped Ge--solid lines (a), (c), and (d)--at 4.2°K and the calculated conductance--dashed line (b)--using the model developed in Ref. 3 for a barrier height $V_b=0.63V$ obtained from capacitance measurements. The contact metal is P_b, and the contact area is $2.5\pm0.5\times10^{-4}cm^2$. LH-252

(OVERLEAF BLANK)

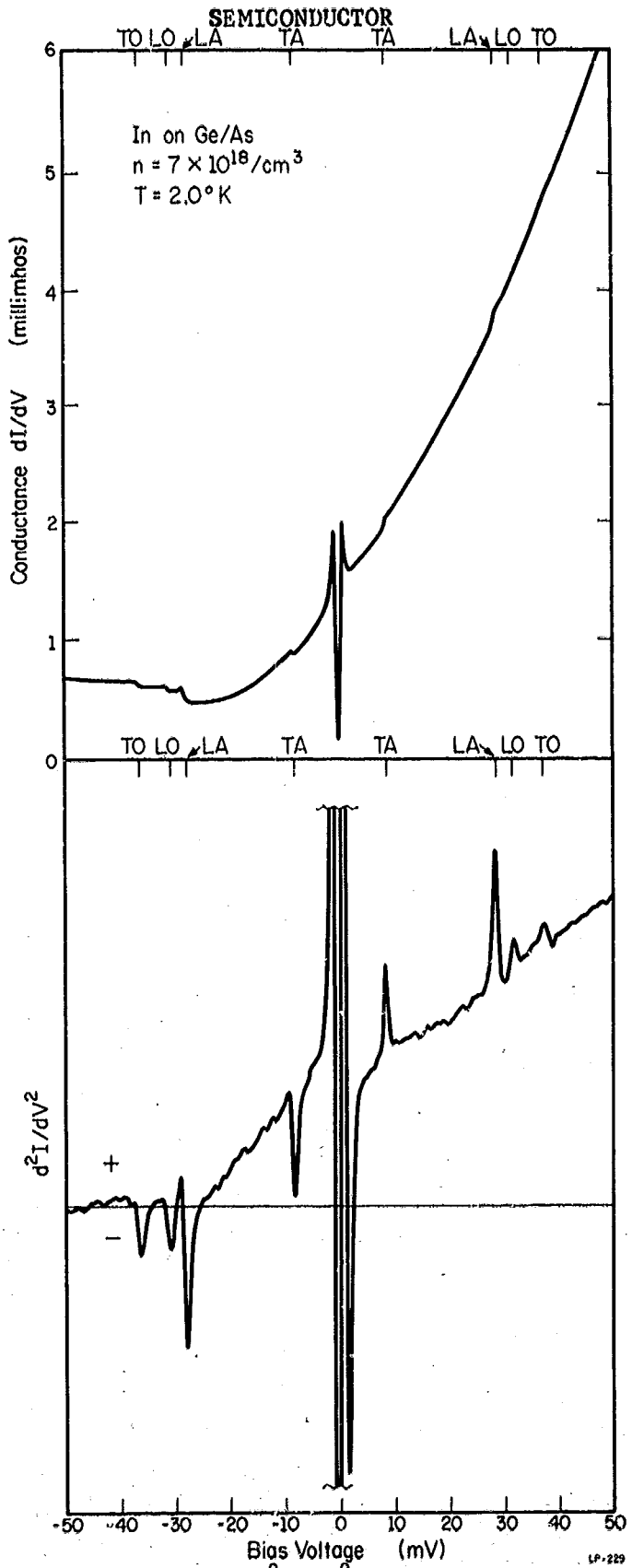


Fig. 7.2. Conductance dI/dV and d^2I/dV^2 of an indium contact on As-doped Ge at 2°K. The doping concentration is $7 \times 10^{18}/cm^3$. The In super-conducting gap at zero bias is shown explicitly. The four zone boundary phonons are shown in both bias directions.

The superconducting energy gap was seen for all junctions (12 Sb- and 10 As-doped samples) below the transition temperature of the contact metal. Pronounced phonon structure was also observed in all units. The influence of ambient gases on the interface of the metal-semiconductor contact was investigated. Two Ge bars were vacuum cleaved, and contacts were evaporated onto the cleaved surfaces. The cleaved-off ends were given a brief exposure to air and then put back into the evaporator where contacts were evaporated onto the air-exposed surfaces. Contacts were made on very closely the same spot on the same bar for both treatments thus tending to insure as nearly the same doping concentration on both interfaces as was possible. The zero-bias conductance of the air-exposed junctions was 7.5 times larger than that of the vacuum-cleaved ones. The contribution to the conductance $\Delta G/G$ by phonon-assisted tunneling was much smaller in air-exposed junctions. The barrier height was found to be 0.51 ± 0.03eV in air-exposed samples. This correlates well with the increased conductance. A similar effect has been found for n GaAs (see below).

7.2.2. Metal-Gallium-Arsenide Junctions*

An attempt was made to study in more detail the phonon line

*We are indebted to Professor N. Holonyak, Jr., University of Illinois and to the Monsanto Company, St. Louis, Mo., for providing us with a variety of GaAs samples. Special thanks are due to Jon Rossi, a student of Prof. Holonyak, for helping us with many aspects of GaAs junction preparation.

shapes reported by Conley and Mahan [5] in metal-gallium-arsenide junctions under conditions in which vacuum cleavage was used to prepare the junction thereby giving a direct comparison with the results on chemically treated samples. Samples with 6.5 and $9.5 \times 10^{18}/cm^3$ Zn-doping were prepared as usual by vacuum cleavage and evaporation of In contacts. The superconducting energy gap was observed in all junctions. In forward bias, strong structure was observed at a bias of about 38 mV which corresponds to the LO phonon. In reverse bias, only weak structure was found at the phonon energy. The structure around zero bias can be explained by hole-TA-phonon interaction. [6] Figure 7.3 shows dI/dV and d^2I/dV^2 for p GaAs with $6.5 \times 10^{18}/cm^3$ Zn doping.

Junctions were made on GaAs doped with 2, 3.85, 6.7, and $9.3 \times 10^{18}/cm^3$ tellurium. Only the two highest doping levels gave useful junction resistances. For lower doping levels, air cleavage was used. This lowered the barrier without changing any structure in d^2I/dV^2. Figure 7.4 shows dI/dV and d^2I/dV^2 for n GaAs ($9.3 \times 10^{18}/cm^3$, Te). Strong, sharp structure is observed at $\pm$ 38 mV due to electron-LO-phonon interaction. The broad weak peak observed at $\pm$ 35 mV is perhaps due to electron-to-phonon interaction. Identical structure was found in Se-doped samples ($5.15 \times 10^{18}/cm^3$).

A strong inflection in d^2I/dV^2 in reverse bias, and a weaker anti-symmetric inflection occurs at forward bias at the same value of bias.

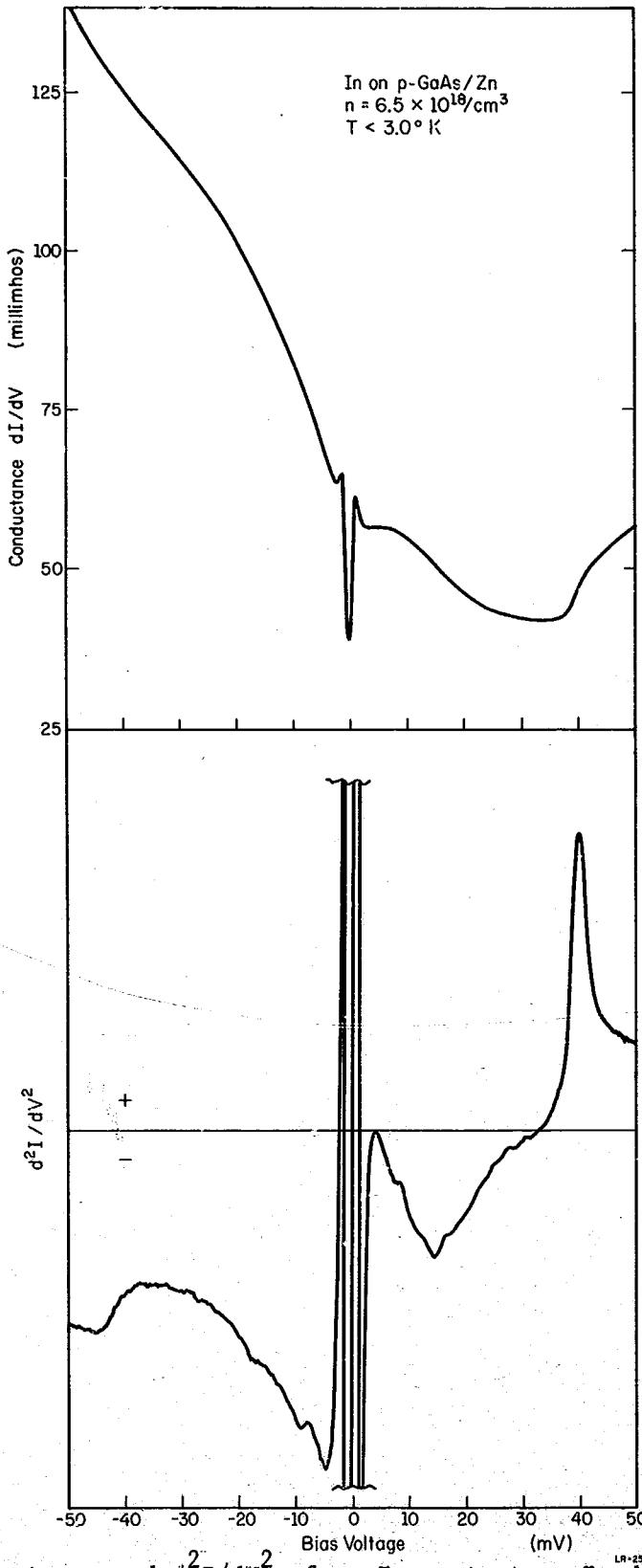


Fig. 7.3. Conductance and d^2I/dV^2 of an In contact on Zn-doped GaAs with $n = 6.5 \times 10^{18}/cm^3$ at $T=3^oK$. The structure at $\pm$ 38 mV is due to the LO phonon. At reverse bias, it is shifted to higher energy due to lead resistance.

(OVERLEAF BLANK)

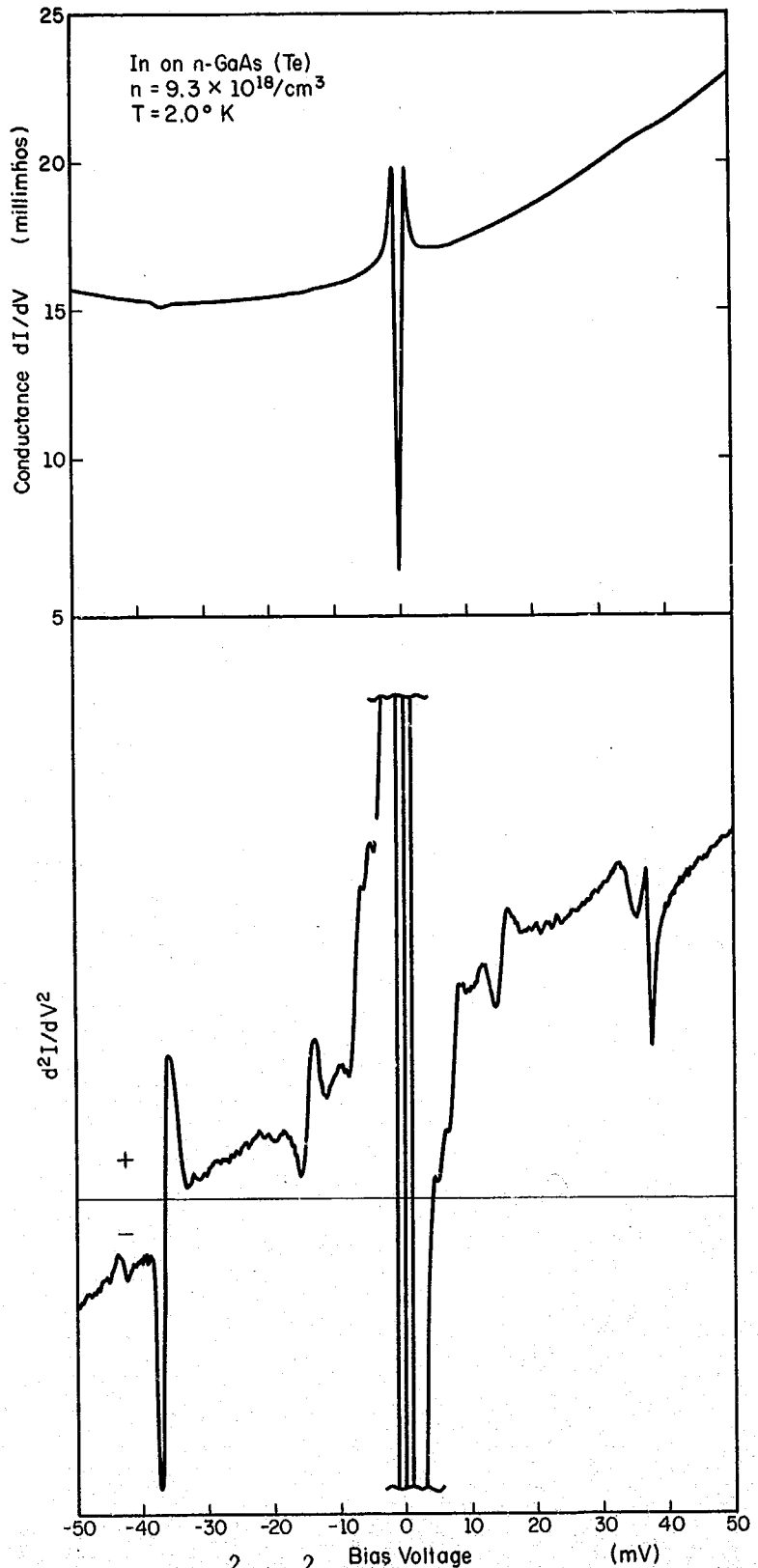


Fig. 7.4. Conductance and d^2I/dV^2 of an In contact on Te-doped GaAs with $n = 9.3 \times 10^{18}/cm^3$ at $T = 2.0^{\circ}K$. The peaks at $\pm$ 38 mV are again due to the LO phonon. The origin of the structure at $\pm$ 35 mV is not yet clear. Structure at 15 mV or less is due to In phonons.

(OVERLEAF BLANK)

These inflections move to higher energies with increasing doping level. This is tentatively interpreted as excitation of the conduction electron plasmon on the part of tunneling electrons. This is another self-energy effect. Plasmons are well known from optical data [7] but have not yet been observed in tunneling measurements. Figure 7.5 shows lineshapes and position vs doping. The discrepancy between observed peakdown position in d^2I/dV^2 (supposedly plasmons) and calculated reflectivity minimum (marked by arrows) could be due to an increase in effective mass with doping concentration as reported in Ref. 7.

7.2.3. Metal-Silicon Junction

An attempt was made to see phonon-assisted tunneling in metal-n-type-Si junctions that was analogous to the tunneling observed in metal-n-type-Ge junctions reported above and to the tunneling in Ge- and Si-p-n junctions. [8] Silicon is quite different from Ge, because the energy difference between direct and indirect conduction band minima is much larger in Si. It also requires a much higher doping concentration for degeneracy (about $2\times10^{19}/cm^3$ compared to about $1\times10^{18}/cm^3$ in Ge).

Tunneling measurements were made with vacuum-cleaved junctions with doping concentrations from 1 to $2.5\times10^{19}/cm^3$ with Sb and P as dopants. Air-cleaved junctions (MIS) were also measured for doping levels up to $1\times10^{20}/cm^3$. Besides the superconducting energy gap of the

contact metal at $T<T_c$, a strong step up in conductance at about 60 mV forward bias was seen in all junctions, including MIS structures. Figure 7.6 shows dI/dV and d^2I/dV^2 for a vacuum-cleaved unit with $1X10^{19}/cm^3$ Sb doping.

A comparison between one-electron theory [3] and experiment shows that the observed conductance (Fig. 7.6) is at least 10^3 times larger than the theoretical value. At a doping level of $2X10^{19}/cm^3$, calculated and observed values agree quite well, however. It is not known at present whether this results from a lowering of the barrier by image forces [9] or some other effect. At $1X10^{19}/cm^3$, one is still in the impurity band range where the theory of Ref. 3 may not necessarily apply. The structure in d^2I/dV^2 that is seen in Fig. 7.6 gradually disappears when the doping level is increased. For higher doping levels, the only surviving strong structure is the peak at about 60 mV. Its origin is not clear, and it behaves completely different from the phonon-assisted peaks in n Ge; e.g., its shape changes very little with temperature, and there is little or no structure at reverse bias. One would also have expected to see at least three phonon peaks. A number of these effects are presently not understood. It is clear, however, that phonon-assisted tunneling in metal-n-type-Si junctions is much less probable than in the comparable Ge system.

Additional structure in the derivative of the conductance, d^2i/dV^2, has been detected in MOS junctions made from degenerate p-type

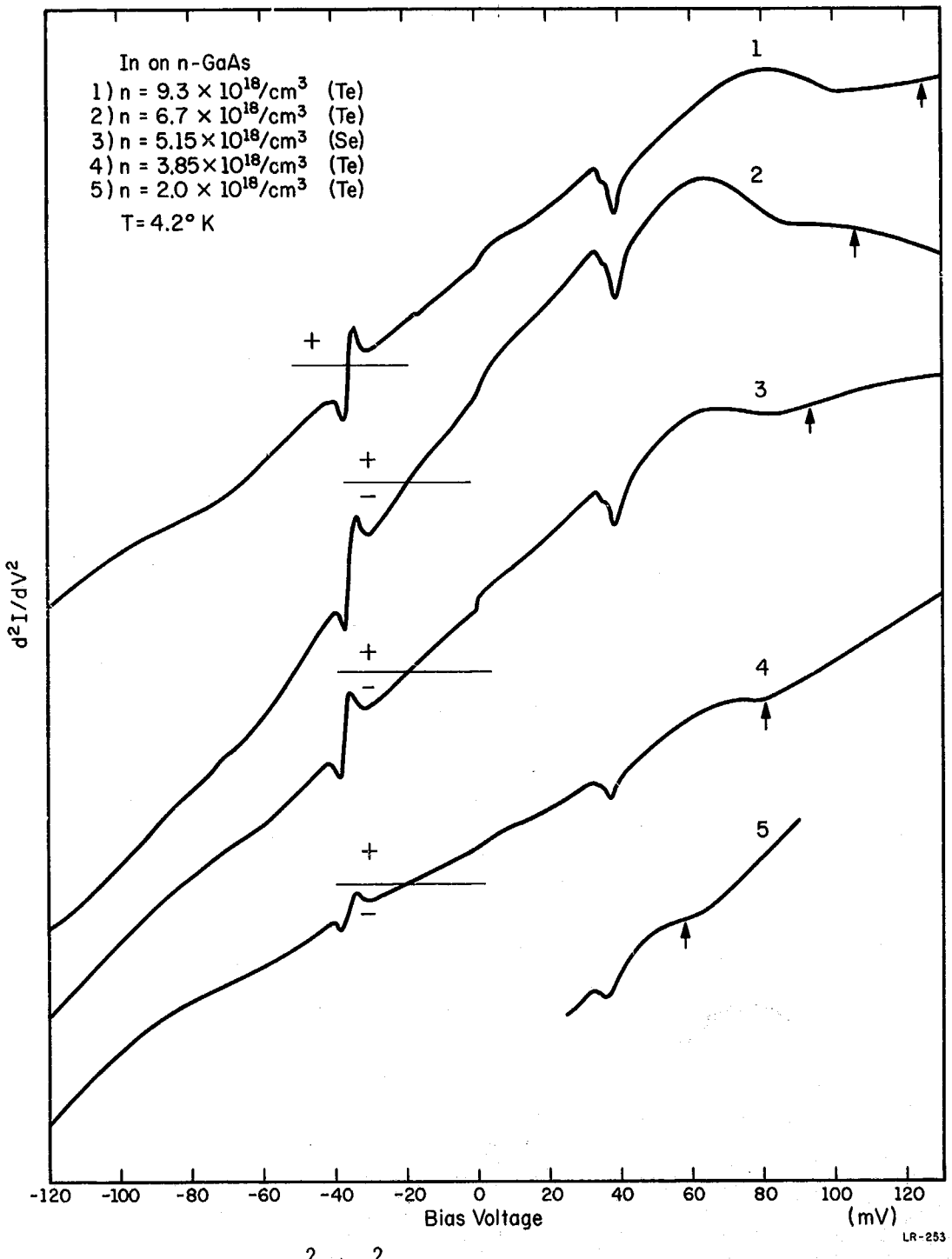


Fig. 7.5. Structure in d^2I/dV^2 in In-n-GaAs junctions of different doping levels at 4.2°K. Arrows show where the Ir-reflectivity minimum as a function of doping concentration should occur under the assumption of a constant electron effective mass.

(OVERLEAF BLANK)

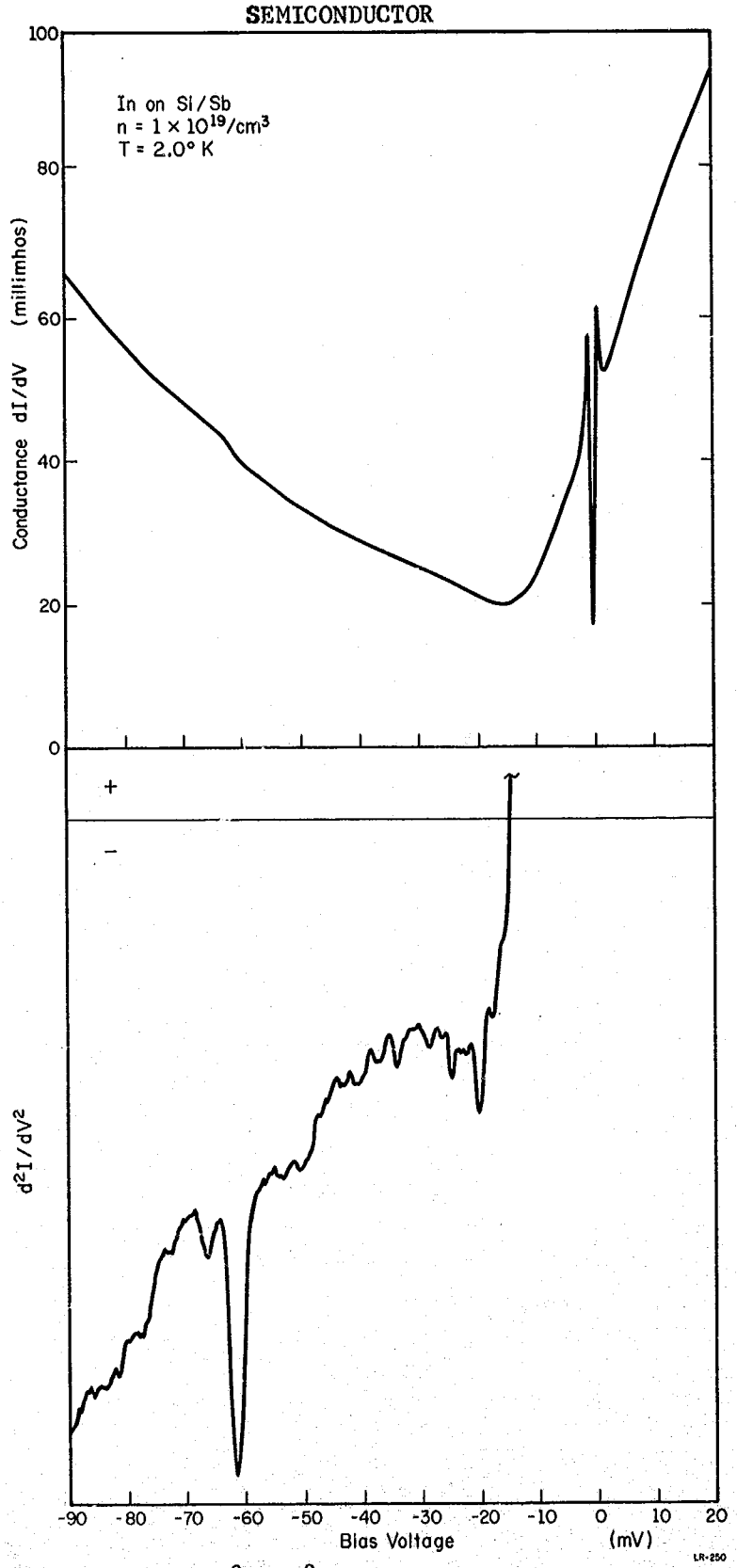


Fig. 7.6. Conductance and d^2I/dV^2 of an In contact on Sb-doped Si with $n = 1 \times 10^{19}/cm^3$ at 2^oK. Due to lead resistance, the peak at 60 mV is shifted to 62 mV. In dI/dV, the superconducting energy gap of In is shown.

(OVERLEAF BLANK)

silicon. (1.3 to about 2.5×10^{20} boron atoms per cm^3) in a manner described on page 140 of Ref. 1. Measurements of d^2i/dV^2 on several of these junctions at temperatures below $4.2^{\circ}K$ have revealed the presence of structure previously undetected. A weak peak at about -77 mV, seen in Fig. 7.7, was thought to arise from the boron impurity. In an effort to positively establish the identity of this peak, a very-highly-doped silicon crystal containing about 2.5×10^{20} boron atoms per cm^3 has been procured. The boron content of this crystal is roughly twice that of any material previously examined. Measurements made on this new crystal have confirmed the earlier thoughts that the structure, see Fig. 7.8, is indeed due to the boron. Other weak peaks, such as that seen in Fig. 7.7 between the boron peak and the silicon phonon structure, have been observed. These peaks are observed at energies roughly corresponding to the sum of the energy of the silicon optical phonon and the energy of one of the lead phonons, and hence it is conjectured that a two-phonon tunneling process may have been observed with one phonon being from the superconductor and the second from the LO-optical phonon of the silicon. This possibility will be investigated by making tunneling measurements in a magnetic field sufficiently strong to quench the superconductivity. The necessary apparatus has been obtained and assembled, and measurements are underway. To date, however, these measurements are inconclusive. The thickness of the superconducting film contact is such as to greatly increase the magnetic

field necessary to quench the superconductivity, thereby producing some uncertainty in identifying when the superconducting energy gap is quenched.

In an effort to make junctions that were more mechanically stable, the fabrication of epoxy junctions as described on page 142 of Ref. 1 was further perfected. Three such junctions on a single slice of silicon are shown in Fig. 7.9. Encouraging results were obtained with some of the first few junctions. They were indeed much more mechanically stable and one junction was cycled to 4.2°K four times with only small changes in the junction resistance. In general, however, the junction resistance of these units proved to be too small for convenient measurement and, at present, junctions made in this manner are not being used.

F. Steinrisser
D. Cullen
L. Schein

7.3. Radiation Damage in Semiconductors

The luminescence arising from electron-hole recombination at radiation damage centers has been found to be a sensitive tool for determining some of the properties of these defects. The energy of the luminescence gives a measure of the location of the energy level in the forbidden gap. The structure of the luminescent-emission band has been shown to arise from phonon interaction with the charges trapped at the defect. [10] Several questions remained unanswered, however. First,

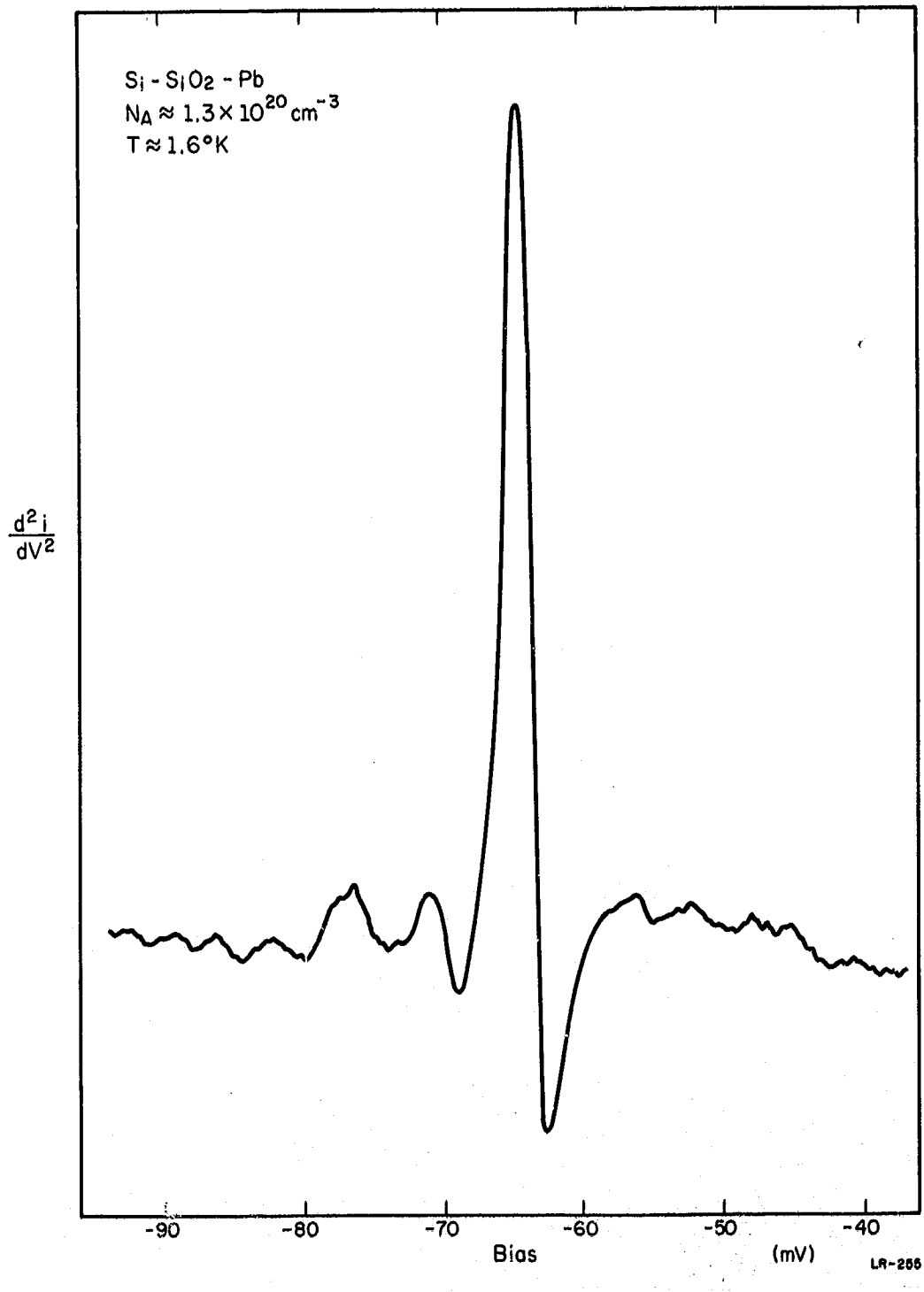


Fig. 7.7. The spectrum of d^2I/dV^2 in the region around negative bias K=0 optical phonon of silicon. The peak at about -77mV is thought to be due to the boron impurity and the remaining peak at about -7/mV is thought to arise from a two-phonon tunneling process involving the optical phonon of silicon and a lead phonon.

(OVERLEAF BLANK)

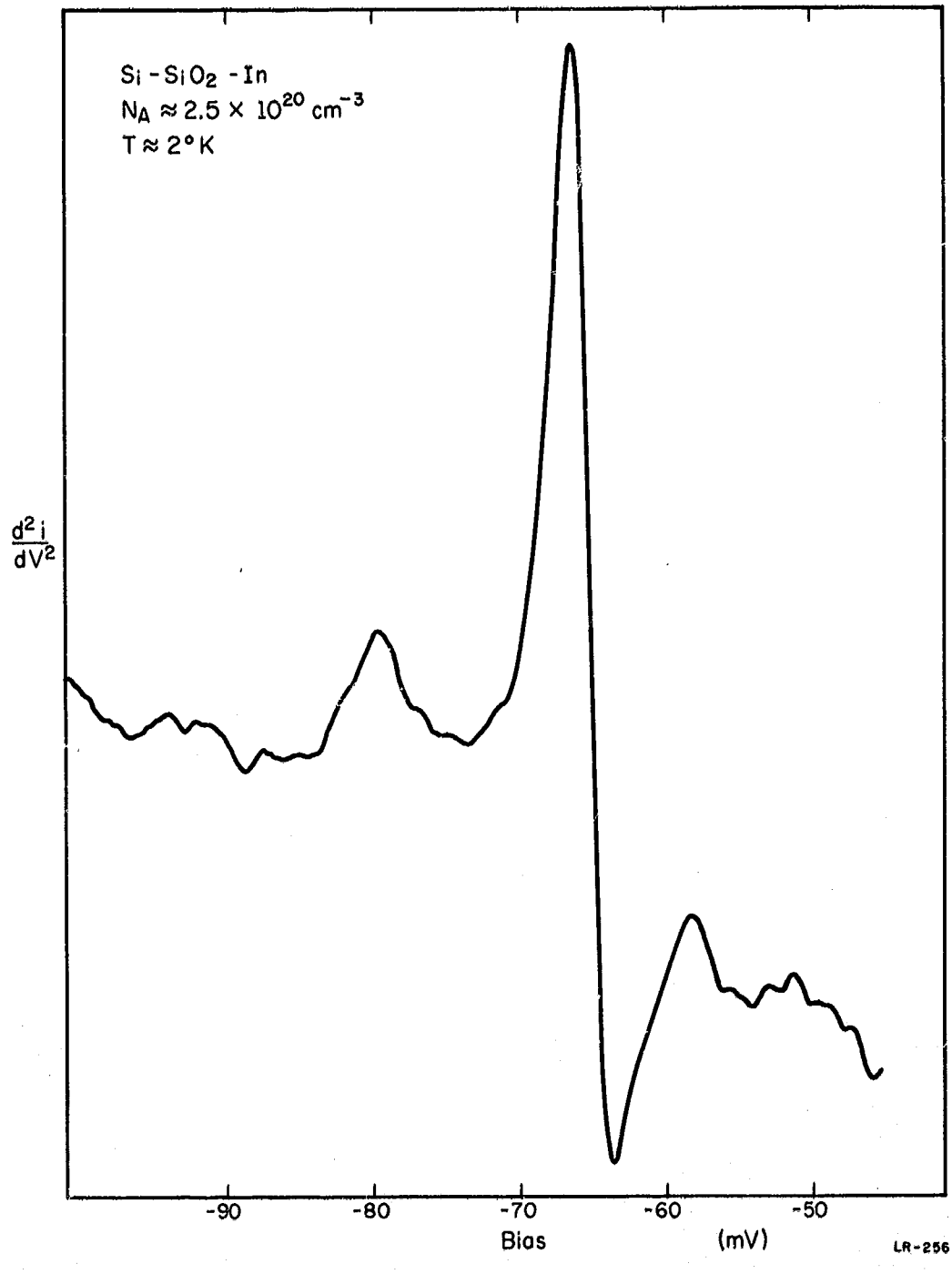


Fig. 7.8. The spectrum of d^2I/dV^2 in the region around negative bias K=0 optical phonon of silicon. The boron local mode is clearly evident in this more highly-doped silicon sample.

(OVERLEAF BLANK)

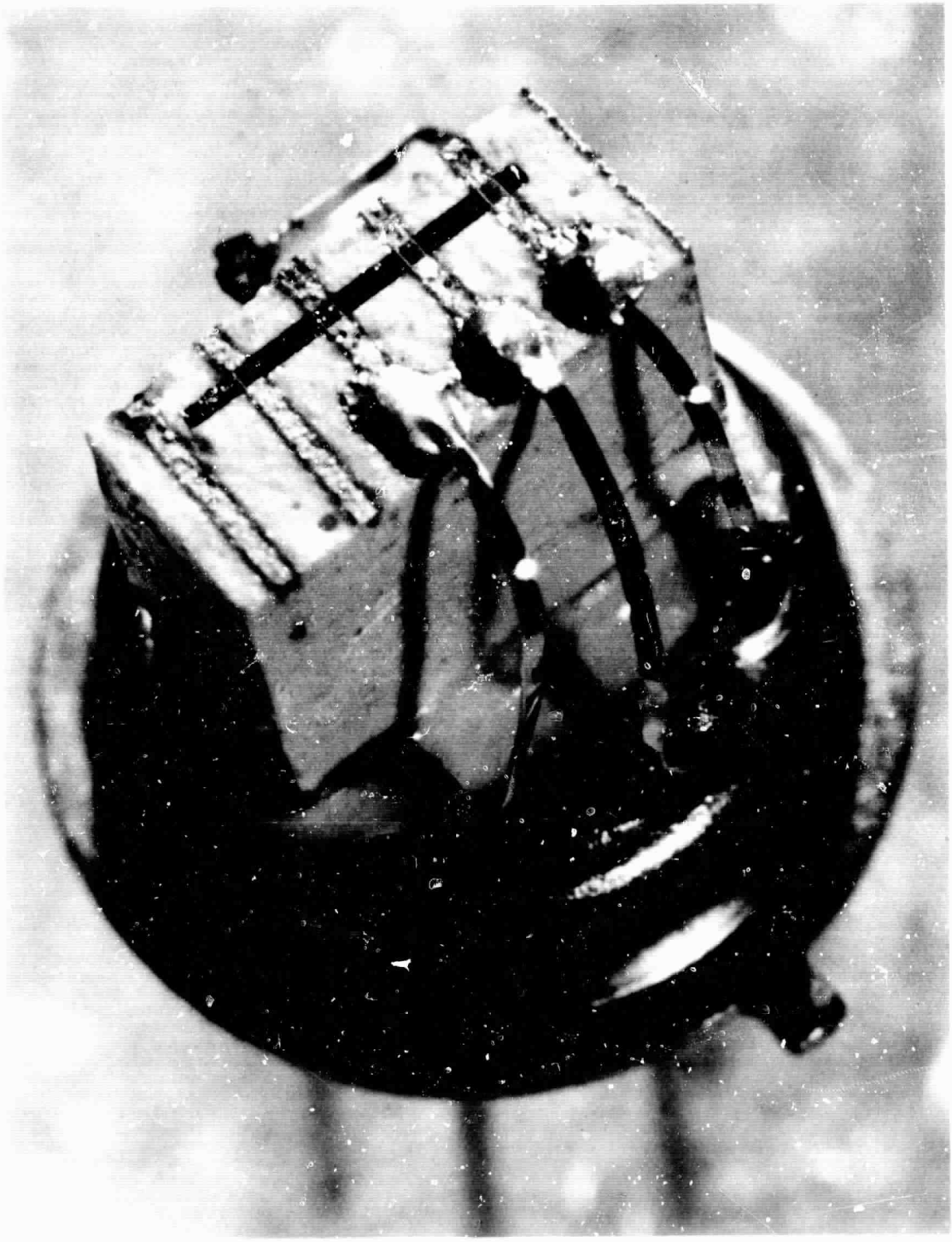


Fig. 7.9. Epoxy-molded silicon slice. Three MS junctions have been made on this unit.

(OVERLEAF BLANK)

were the transitions that gave rise to the luminescence those between bound electrons and bound holes or between bound electrons and free holes? Second, what is the role of oxygen in the formation of the particular defects that appear? Third, can the thermal annealing of the luminescence be used as a means of correlating these results with other measurements? Fourth, what is the influence of impurities such as lithium upon the luminescence?

Careful measurements of the temperature dependence of the half width of the zero-phonon lines are underway using the equipment described in Ref. 1. The line at 1.278μm has been observed at temperatures from 6°K to 40°K. The data thus far have a large scatter due to changes in the background luminescence. It is possible to say that the observed half width is less than 0.6 nm at 6°K and less than 1.0 nm at 40°K with an instrument resolution of 0.36 nm. If the line shape is approximated by a gaussian, the actual half width is related to the observed half width by [11]

$$\Delta_{\frac{1}{2}act}{}^2 = \Delta_{\frac{1}{2}obs}{}^2 - \Delta_{\frac{1}{2}ins}{}^2 ,$$

where $\Delta_{\frac{1}{2}obs}$ is the half width of the line observed with an instrumental resolution of $\Delta_{\frac{1}{2}ins}$, and $\Delta_{\frac{1}{2}act}$ is the actual half width of the line. Using this expression, we find a half width of 0.5 ± 0.1 nm at 6°K and 0.8 ± 0.1 nm at 40°K.

The data for the peak at 1.57μm is better, since the intensity is stronger and the background did not vary in the vicinity of the peak. Using the above formula, the half width of the line appeared to increase slightly as the temperature is increased from $6^{o}K$ to $30^{o}K$, being 0.32 ± 0.10 nm at $6^{o}K$ and 0.4 ± 0.10 nm at $30^{o}K$. The width appears to increase rapidly above $30^{o}K$.

The total intensity as measured with broad spectral resolution, of the two zero-phonon lines increases as the temperature is raised from $4.2^{o}K$ to about $30^{o}K$. Above $30^{o}K$ the intensity decreases sharply. A plot of the logarithm of the integrated intensity vs 1/kT for the low-temperature region yields a straight line with a slope of approximately one milli-electron volt. A similar plot for the high-temperature region also yields a straight line with a slope of between 10 and 20 millivolts for both the 1.278μm and 1.57μm bands. It is expected that a more careful measurement of the activation energy for quenching the luminescence will give information concerning the binding of at least one of the carriers to the defect.

Recombination luminescence is quenched in pulled silicon by the diffusion of lithium. Neither band-to-band nor radiation-induced defect luminescence has been seen in this material to date. The lithium concentration was on the order of $10^{18}/cm^{3}$. Luminescence has been seen in lithium-drifted, float zone material but is of low intensity. After

irradiation with fast electrons of 10^{16} and 10^{17} e/cm^2 fluxes, the recombination luminescence consists of a broad band located between 1.00 and 0.78 eV which is essentially structureless at these low intensities. It is not clear at this time whether this luminescence has any relationship to the band at 0.97 eV previously reported by Spry and Compton.[10]

Samples of 100 Ω cm N-type pulled silicon were irradiated with 10^{15}, 10^{16} and 10^{17} fast electrons/cm^2. An intense luminescence peak is found at 0.97 eV that differs by at most a factor of two between the 10^{15} and 10^{17} irradiations. A band at 0.79 eV is barely discernible. After standing approximately two weeks at room temperature, the intensity of the 0.79 eV band grows to about one half of that of the 0.97 eV band. Upon remaining longer at room temperature, the 0.79 eV band decays until it is barely discernible after 5 weeks. This annealing pattern appears to be the same as observed by Spry for gamma irradiated material. [10] A similar annealing behavior seems to exist in samples kept at dry-ice temperatures.

The spectrum of a 100 Ω cm N-type pulled sample irradiated with 10^{16} e/cm^2 is shown after a room temperature anneal of 53 days in Fig. 7.10. Figure 7.11 presents the spectrum that exists after a 14 hr anneal at 300°C. Spectacular changes have occurred in the spectrum. The 0.97-eV luminescence peak has disappeared completely leaving a new series commencing at 0.95 eV. The peak separation among the resolvable peaks in this family of

bands is identical to that of the family commencing at 0.97 eV in freshly irradiated material. The intensity of the 0.79 eV peak and its associated family of peaks has been greatly enhanced by the anneal. The position of the peaks in this family are the same as previously observed. [10]

In the work of Spry, [10] a weak band was observed on one occasion some 0.005 eV on the high energy side of the 0.79 eV peak in unannealed material. After the 300°C anneal, a band appears in this position whose intensity is second only to the 0.79 eV line itself. Further detailed studies of these phenomena are underway.

7.4. Investigation of p-i-n Diodes as Infrared Detectors

The investigation of p-i-n diodes as infrared detectors has continued with detailed measurements being made on a number of gold-doped silicon diodes. The frequency of oscillation at threshold, defined as the frequency of the first detectable oscillation that occurs as the bias voltage is increased from zero, was measured as a function of temperature. Although oscillatory effects are found for various dopants in silicon, e.g., Co, Zn, and Au, the Au-doped units are particularly interesting because of their high frequency, thus offering the possibility of fast response to an external infrared signal. [12]

Figure 7.12 is an illustration of the test setup, including bias circuit, environmental chamber, and measurement apparatus. Threshold frequencies were observed over a range of 2 to 100 MHz. Because of the

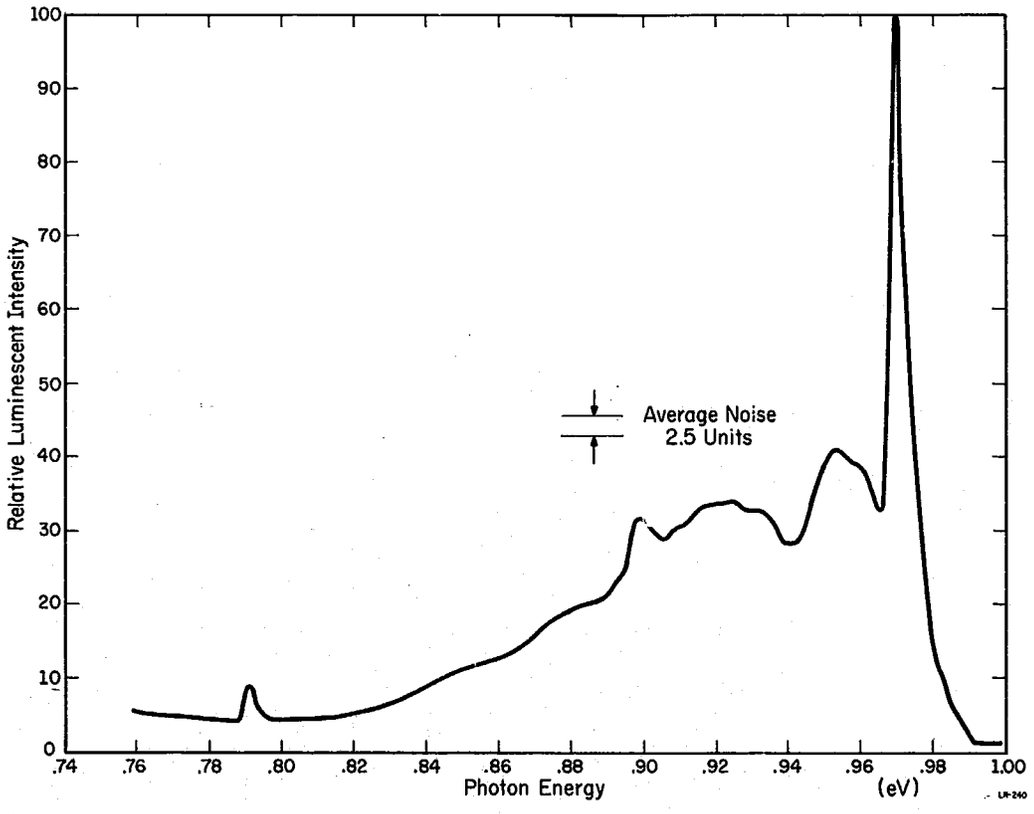


Fig. 7.10. Recombination luminescence of 100 Ωcm N-type silicon following irradiation with 10^{16}e/cm^2. Sample remained 53 days at room temperature prior to measurement at liquid-helium temperature. The slit width is 1 mm.

(OVERLEAF BLANK)

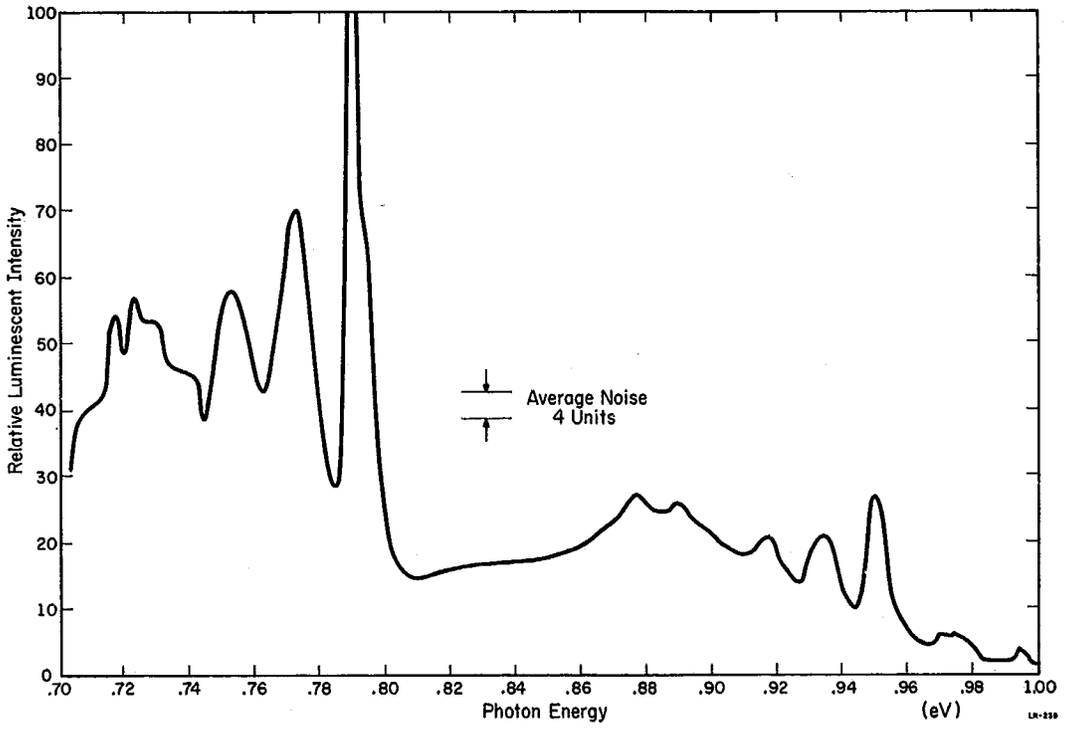


Fig. 7.11. Recombination luminescence of 100 Ωcm N type silicon following irradiation with 10^{16}e/cm^2. Sample was annealed for 14 hours at 300°C after measurement shown in Fig. 7.10. Measured at liquid-helium temperature. The slit width is 1 mm.

(OVERLEAF BLANK)

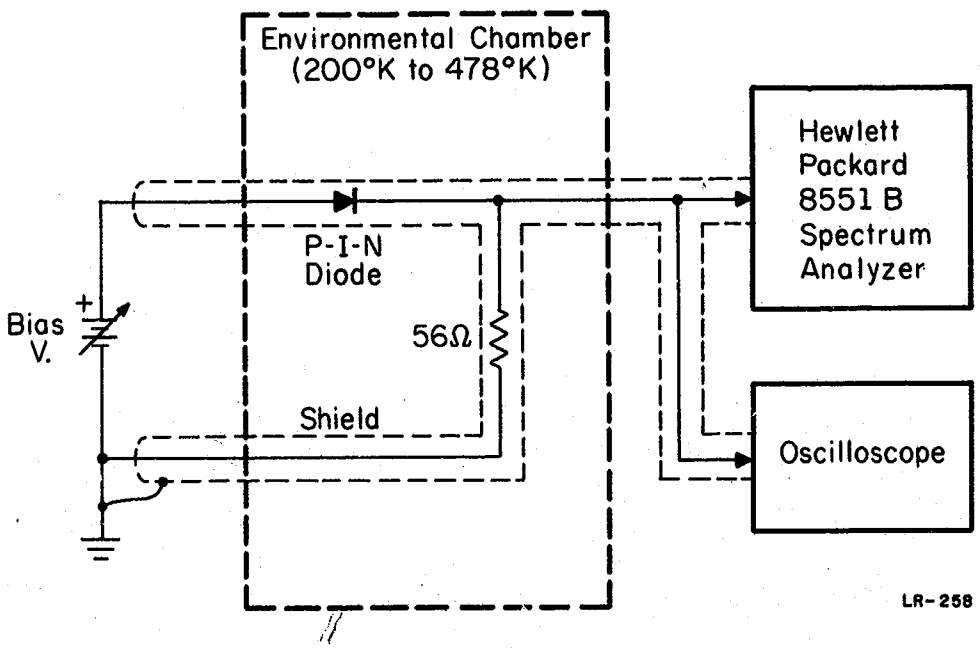


Fig. 7.12. Experimental arrangement for determining the threshold oscillation frequency of p-i-n diodes.

(OVERLEAF BLANK)

high oscillation frequencies, the diode samples and the load resistor were placed inside shielded holders. The background radiation temperature was therefore the same as the device temperature.

Figure 7.13 presents data from four diodes on the threshold frequency *vs* reciprocal temperature. In general, the plots exhibit a straight line characteristic indicating an exponential relationship of the form $f=f_o e^{-Q/KT}$. The value of Q obtained graphically from the slope of the curves and the threshold oscillation frequency at 300°K is given in Table 7.1, for the various diodes tested. Information pertaining to the fabrication of the diodes is also included in the table. Two diodes, as shown in the table and on the curves, exhibit two distinct values of slope.

Table 7.1
Characteristics of a number of Au-doped p-i-n silicon diodes

Sample	MX1	BS Au2	MC1	MZ1	MB2	MB1	MY1	M Au1
Base Material of Diode	n	n	p	n	n	n	n	n
Temp. of Au diffusion (deg. C)	900	900	987	1025	1025	1025	1100	1100
Diffusion Time	2½ days	6 days	11 hours	23 hours	23 hours	23 hours	46 hours	46 hours
a(eV±0.01)	0.14	0.18	$\frac{0.11}{0.24}$	0.13	$\frac{0.16}{0.28}$	0.17	0.18	0.24
Threshold Oscillation Frequency at 300°K (MHz)	9.4	10	15.3	36	33	32	53	35
Au content as determined by diffusion temp. [12]			9×10^{15}	1.3×10^{16}	1.3×10^{16}	1.3×10^{16}	3×10^{16}	3×10^{16}

The gold content in these diodes was determined from the gold solubility data of Trumbore. [12] A comparison of the measured threshold frequencies *vs* Au concentration was made with the data of Moore, *et al.* [13] The threshold frequencies for the high-temperature diffusions is substantially lower than reported by Moore for comparable diffusion temperatures.

The origin of an activation energy near 0.2 eV for the Au levels is not presently understood. This does not correlate with any of the common energy levels of gold in silicon. In addition to the gold-doped units, data were also taken of oscillation frequency *vs* temperature for a cobalt-doped diode and the detectivity (D^*) versus temperature for a zinc-doped sample. The temperature dependence of the oscillation frequency for a cobalt-doped silicon diode is shown in Fig. 7.14. These curves also exhibit more than one value of slope. The values of Q for each of these slopes are shown on the curves. In this case, two radiation background temperatures were used, one at environment background, and one at 1373^{o}K. The upper curve clearly demonstrates the effects of infrared radiation upon the frequency response. The measured energy level of about 0.30 eV corresponds reasonably well with the known level of 0.35 eV of Co in silicon. [14]

The detectivity curve for a zinc-doped silicon diode is shown in Fig. 7.15. The detectivity was measured in the photo-conductive mode. The curve shows the detectivity to be increasing very rapidly as the diode is cooled. It is to be noted that even at room temperature the detectivity is comparable to state-of-the-art detectors and improves as the diode is cooled.

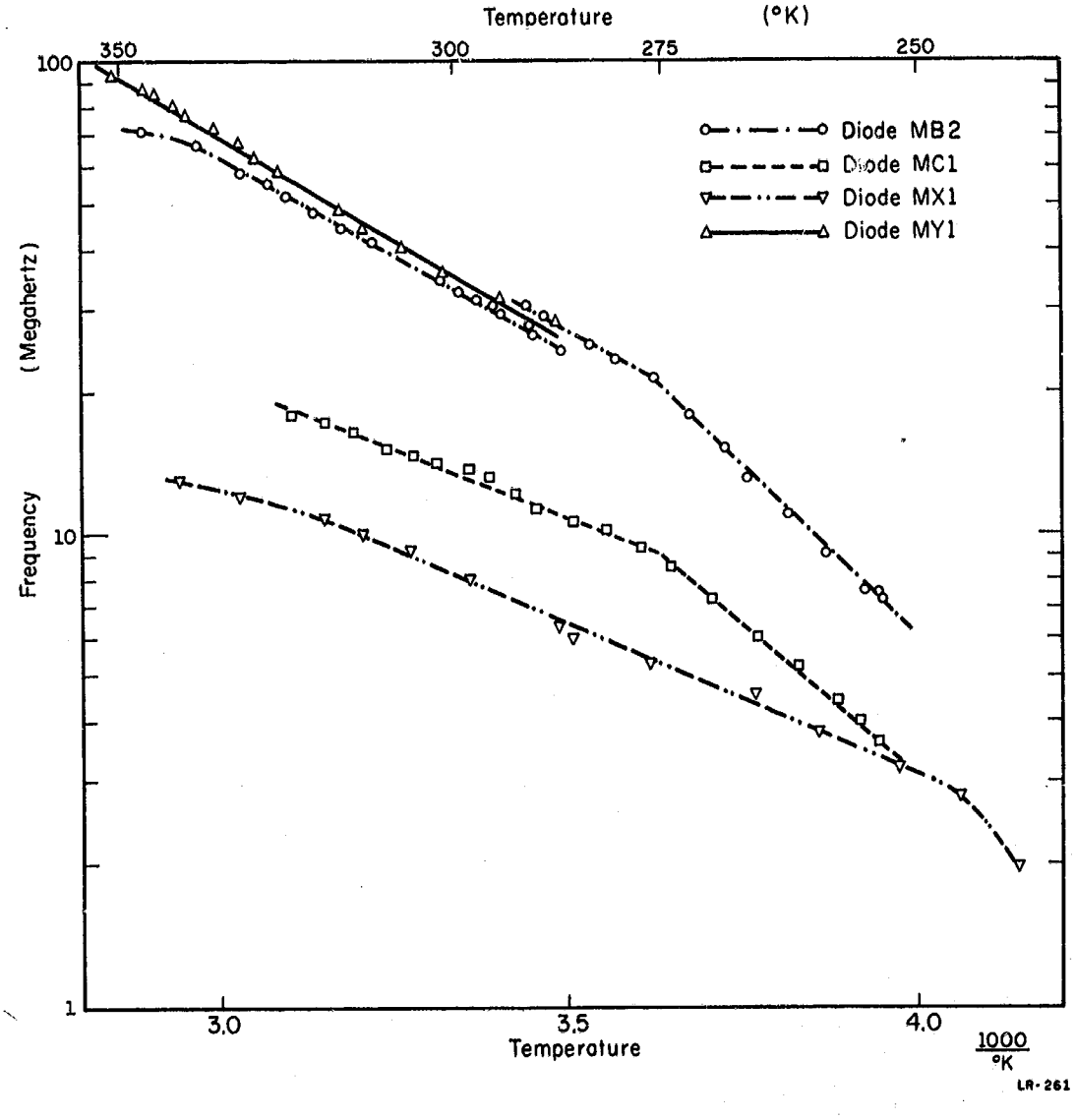


Fig. 7.13. Threshold oscillation frequency vs reciprocal temperature for four gold-doped p-i-n silicon diodes. Characteristics of the units are given in Table 7.1.

(OVERLEAF BLANK)

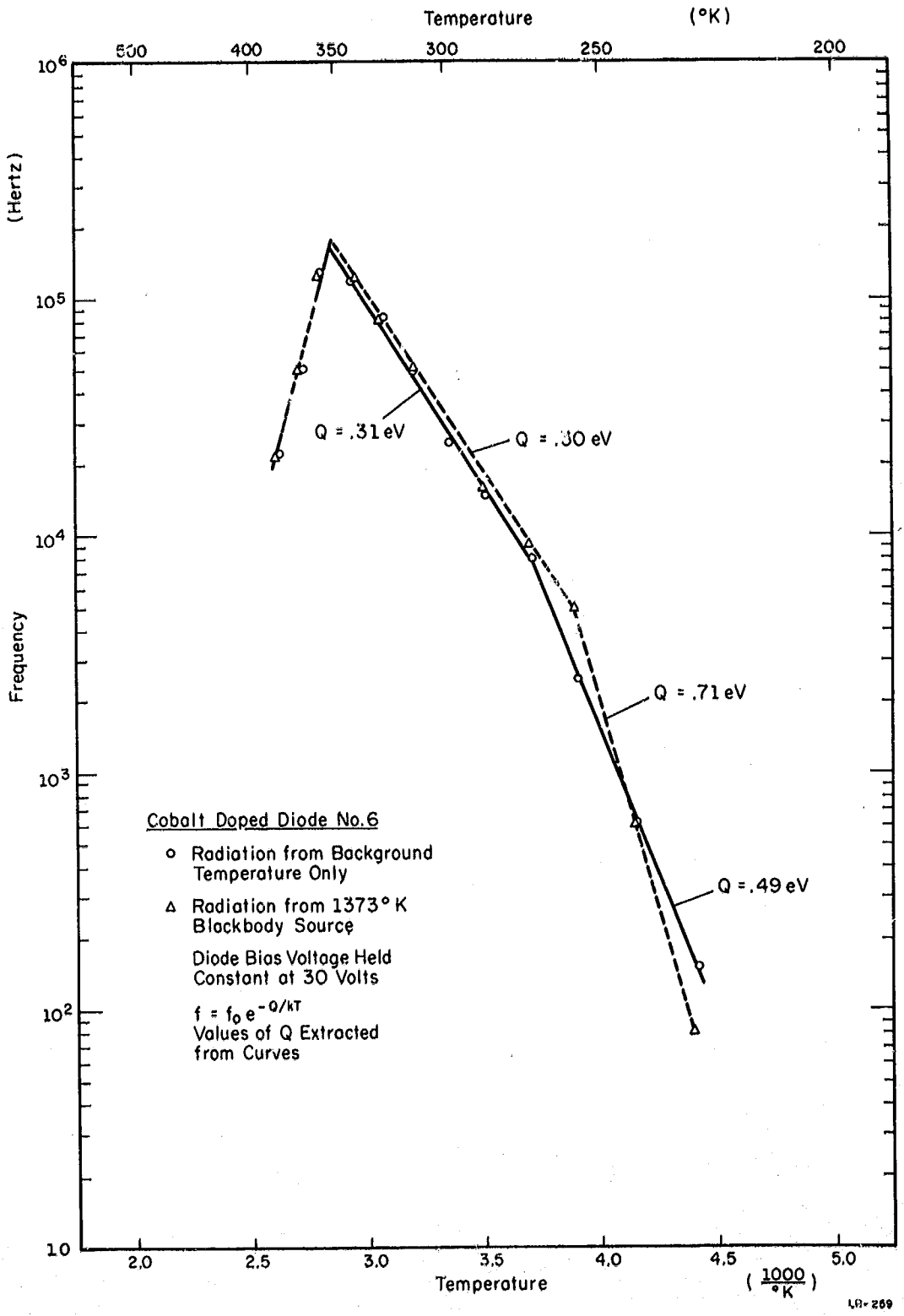


Fig. 7.14. Oscillation frequency vs reciprocal temperature for a cobalt-doped p-i-n silicon diode. Upper curve--diode irradiated with a 1373°K blackbody source. Lower curve--diode irradiated with 300°K blackbody source.

(OVERLEAF BLANK)

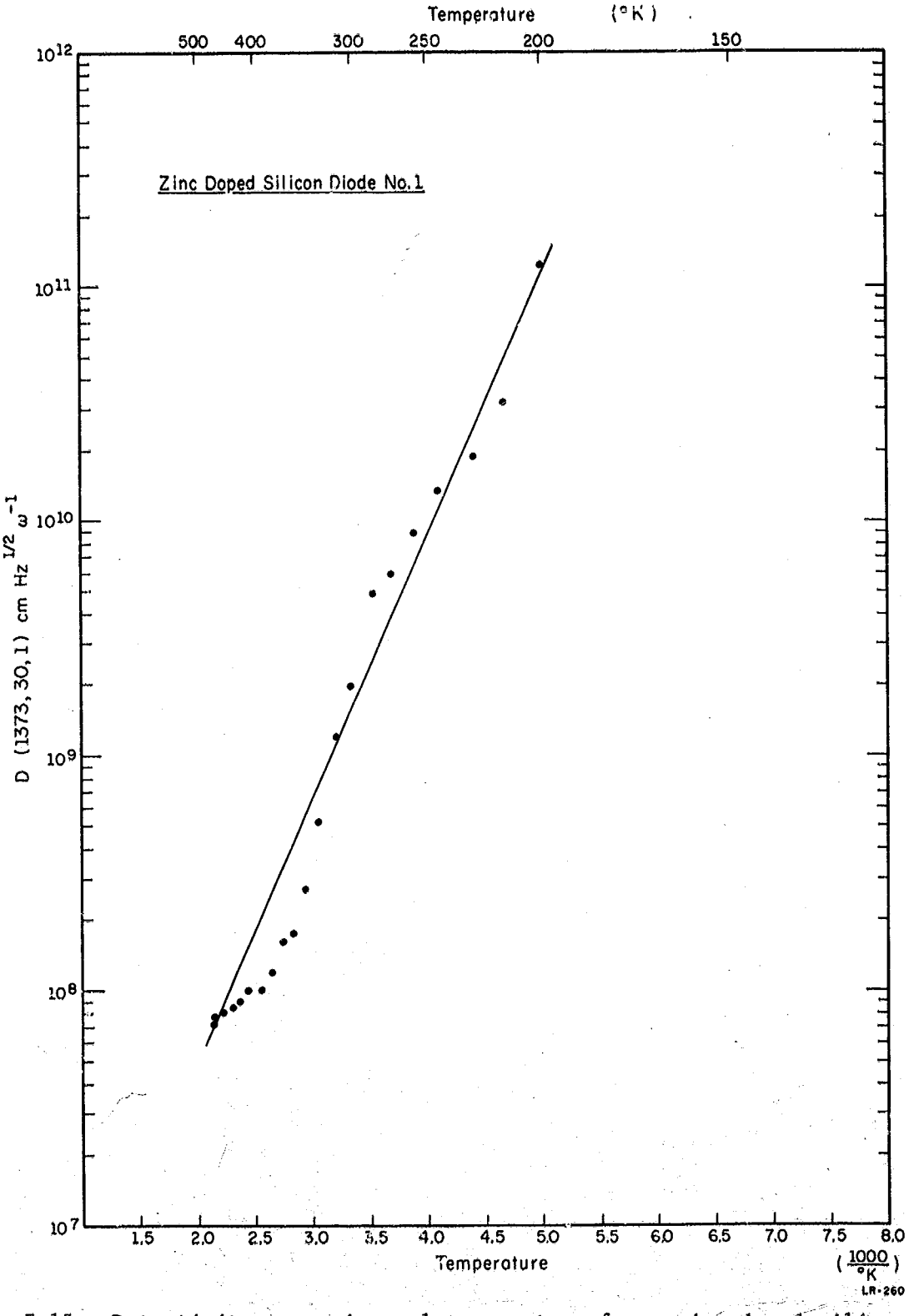


Fig. 7.15. Detectivity vs reciprocal temperature for a zinc-doped silicon p-i-n diode.

(OVERLEAF BLANK)

7.5. References

1. Coordinated Science Laboratory Progress Report, September 1967 - February 1968.

2. "Electron and Phonon Tunneling Spectroscopy in Metal-Germanium Contacts", by F. Steinrisser, L. C. Davis, and C. B. Duke (To be published in Phys. Rev.).

3. J. W. Conley, C. B. Duke, G. D. Mahan, and J. J. Tiemann, Phys. Rev. 150, 466 (1966).

4. W. G. Spitzer, F. A. Trumbore, and R. A. Logan, J. Appl. Phys. 32, 1822 (1961).

5. J. W. Conley and G. D. Mahan, Phys. Rev. 161, 681 (1967).

6. A. J. Bennett, C. B. Duke, and S. D. Silverstein, to be published.

7. W. G. Spitzer and J. M. Whelan, Phys. Rev. 114, 59 (1959).

8. A. G. Chynoweth, R. A. Logan, and D. E. Thomas, Phys. Rev. 125, 877 (1962).

9. This mechanism was suggested by C. B. Duke.

10. "Recombination Luminescence of Irradiated Silicon" by R. J. Spry and W. D. Compton in Radiation Effects in Semiconductors, p. 421, F. L. Vook, Ed., Plenum Press, New York (1968).

11. K. S. Seshadri and R. N. Jones, Spectrochimica Acta 19, 1013 (1963).

12. F. A. Trumbore, Bell System Tech. J. 39, 205 (1960).

13. J. S. Moore, N. Holonyak and M. D. Sirkis, Solid-State Electronics 10, 823 (1967).

14. C. M. Penchina, J. S. Moore and N. Holonyak, Jr., Phys. Rev. 143, 634 (1966).

(OVERLEAF BLANK)

8. COMPUTER[†]

H. G. Slottow
J. Bouknight
G. Crawford
L. Hedges
J. Knoke
D. Lee
V. Metze
E. Neff
R. Resch
J. Stifle
R. Trogdon

8.1. Introduction

The Computer Group is responsible for the management, maintenance, and development of the CSL computer facility and for the design of custom digital equipment to meet the needs of the Laboratory.

8.2. CDC 1604 Computer

Period: March 1, 1968 to August 29, 1968

Total Running Time:		3039.00 hours
Average Per Day (7 day week)		16.69 hours
Operational Time:	97.40%	2960.00 hours
Preventive Maintenance Time:	2.00%	61.00 hours
Emergency Maintenance Time:	0.60%	18.00 hours

E. Neff

8.3. New 16mm Motion-Picture Camera System

During the reporting period, a Mitchel 16mm motion-picture camera was acquired by the Computer Group for the purpose of upgrading the computer animation facility at CSL.

The drive motor for the camera was replaced by a stop-action computer-controlled drive system built by the Lafayette Instrument Company

[†]This work was supported by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

of Lafayette, Indiana. A movable mount for the camera was placed under construction to allow the camera to be placed on or removed from the CSL graphics console.

The graphics console will be redesigned in the camera-control section to allow the following control operations:

Mode 0 Control Word	Function
00000	Select Main Scope
00001	Select Camera Scope and Still Camera
00002	Select Camera Scope and Motion Picture Camera
00003	Open Shutter
00004	Close Shutter
00005	Advance Film 1 Frame
00006	Reverse Film 1 Frame

Software production for the camera system will begin in the next reporting period. Emphasis will be placed on a system which is driven by a natural language and allows as complete a computer control of the movie production as is possible.

Two types of projection systems will be available: 1) standard: one projector, and 2) four projectors placed orthogonally with respect to each other for 360-degree projection. Color will be available for both systems.

J. Bouknight
J. Stifle

8.4. Further Development of CSLx

During the reporting period, checkout of the initial version of CSL6 was completed. An alternate version, CSL7, was also incorporated which provides a different storage allocation method.

A preliminary reference manual was also issued during the reporting period. The manual is for reference purposes only and therefore, available for internal distribution only.

Several new statements which do not appear in the original L^6 from Bell Labs [1] have been added to enhance the generality of the CSLx languages. These statements will be detailed in a descriptive system report during the next reporting period.

J. Bouknight

8.5. Fast Fourier Transforms

A program specifically designed for efficient digital analysis of spectra, the Fast-Fourier-Transform program (FFT), has been obtained from Horlick of the Chemistry Department of the University of Illinois and adapted for use at the Coordinated Science Laboratory. The FFT algorithm was originally developed by Cooley and Tukey of Bell Laboratories [1]. It is now widely used for Fourier analysis because of its high calculation speed. At CSL it has been primarily used for analysis of seismic data taken by the Seismic Detection Group.

The "core" of the program--that part which actually computes the transform--has remained unchanged [2]. However, the program has otherwise

been extensively modified to allow the input of data in several different formats and to permit on-line, conversational operation. Two main versions of the program exist at the Laboratory at present, tailored to the needs of the current users. (Other versions of the program can, of course, be easily prepared to handle data in any format.) One of the present versions allows for the input of up to 8192 input points and the output of a spectrum of 4096 points. Input to this program may be from paper tape punched in flex code or in a special format used by Skaperdas for recording data taken by a Michelson far-infrared interferometer or from a BCD magnetic tape previously prepared by another program, STBITS [3], which interprets a data tape generated by the DSE tape unit used by the Seismic Detection Group for producing a magnetic tape with digital information from an analog input tape. The second basic version of the FFT program allows input of 32,768 points of 12-bit accuracy directly from an analog to digital converter which is on-line to the 1604. (The program could be set up to accept 65,536 points of six-bit accuracy.) It appears that the use of the on-line analog-to-digital converter is a more efficient use of both human and computer time than the three-step method involving the DSE tape unit.

An attempt has been made to provide flexible on-line operation, allowing the user to select which subset of the input points he wishes to display and Fourier analyze. (If he does not specify a segment, the data is displayed in four blocks of 8192 points.) The selected input data is first displayed on an oscilloscope connected to the 1604. The user may then decide either to compute the transform or to look at another section of the

input data. If he decides in favor of the latter course he may, if he desires, once again select the number of points and the segment of data in which he is interested. This series of operations may be repeated over and over again, with the user selecting any subset of the data and having the transform calculated and displayed. Whenever the transform is displayed, the user is permitted to request the display of only those segments in which he is particularly interested. The selected points are displayed on an expanded scale. (The expansion is in the "x" direction only; the "y" scale remains unchanged.)

It is possible to add headings to the oscilloscope display of either the input data or the spectrum, and to take Polaroid photographs of any of the displays.

The program may be expanded in the future to allow the plotting of multiple spectra on the same graph. It is also possible that the program may be modified to allow use of the light pen for program control and selection of data subsets. Because the data tapes used by the Seismic Detection Group are in a fixed format, the full flexibility of the STBITS program is not necessary, and a subroutine to directly read the DSE tape may be inserted in their FFT program, thus eliminating one step in the three-step process now necessary when the DSE unit is used.

V. Metze

8.6. Seminars

Members of the Computer Group held several teaching sessions this summer to familiarize laboratory programmers with the new systems which are now available on the 1604.

The first series of seminars on the ILLAR system was conducted, with Karl Kelley giving an overview of the system and discussing the capabilities of its sophisticated assembler, ILLAR. Virginia Metze gave a summary of the features of the system's FORTRAN, comparing it with the older FORTRAN '60 and with CDC FORTRAN '63, and presenting new statements which are a "superset" which has been added to the language at CSL. Charles Arnold explained usage of DISPLAY, a flexible, general purpose oscilloscope output routine which may be controlled through FORTRAN FORMAT statements and special statements embedded in the CSL FORTRAN compiler. The ILLAR seminar series will be repeated in the fall.

A series of informal seminars on the CSL6-CSL7 list processing language was given by Jack Bouknight, who implemented the language here. Virginia Metze conducted two FORTRAN classes, one on a beginning level, and one on an advanced level. Both plan to have similar sessions in the fall.

8.7. References

1. J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Math of Comput. 19, 297 (April, 1965).

2. More information about this portion of the program may be found in a memo entitled "Digital Processing of Interferograms," prepared by A. Zachor of Bell Telephone Laboratories and and dated May 24, 1966.

3. STBITS is a program especially written to read magnetic tapes prepared by the DSE tape unit. It can presently input a magnetic tape record of up to 2048 48-bit words in length and break up these words into appropriate 6, 12, or 14-bit segments. It must distinguish between three kinds of information: header bits, digital scan, and data scan. The user specifies the format of the input tape; for example, how many heading bits, how many channels of data scan input, and how many channels of digital scan input. The program also automatically sets up the appropriate format for the printer output and produces a standard BCD data tape for input to the FFT program.

(OVERLEAF BLANK)

9. PLASMA DISPLAY[†]

D. L. Bitzer
H. G. Slottow
B. Arora
W. Goede
R. Johnson
W. Petty
J. Schermerhorn
R. Trogdon

9.1. Introduction

The invention of the Plasma Display Panel at the Coordinated Science Laboratory was a response to the anticipated display needs of the PLATO computer-based Education System. Responding directly to digital information, it stores and displays this information on a rectangular array of bistable gas discharge cells. The essential structure is simple, and with development, it should be inexpensive to produce.

The principles of the Plasma Display have been described in several papers, and in preceding CSL progress reports. In this report, we describe a procedure for obtaining variable intensity; we present a theory for the stability of firing times in the Plasma Display cell; and we describe a new approach to circuit design in which the functions of addressing and decoding are combined in a way that reduces the required number of active elements.

9.2. A Procedure for Obtaining Variable Intensity in the Plasma Display Panel

The effective extension of the Plasma Display technique to provide variable intensity, or grey scale, implies the existence of two properties. First, a means must exist by which the intensity of a cell can be varied. Second, it must be possible to sustain different cells in

[†]This work is supported by the Advanced Research Projects Agency and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

an array at different intensities. The first property, by itself, would make possible only a way of controlling the overall brightness of an image on the Plasma Display Panel. The second property, however, makes possible the shading that is characteristic of images produced by half-tone printing or observed in television.

Several procedures for obtaining grey scale have already been described. One of these replaces each cell by a group of cells from which the light is selectively filtered. With n cells in each group it is possible to obtain 2^n brightness levels. A second procedure allows control of the number of discharges that occur in a basic time interval--a TV field time, for example. Unlike the cell-group process, this process does not increase the number of required cells. It is, however, restricted to applications where the image is changed, or is renewed, after each basic interval.

We describe, here, a new process in which the intensity of each cell can be set independently and can be sustained indefinitely just as the single intensity is now sustained. This approach depends on two properties of the Plasma Display cell. First, the intensity of a discharge is related to the form of the exciting waveform at the time of firing and to the time during which charged particles are collected. For some cells, this relation can be expressed as a function of the slope of the exciting signal. Second, with a properly-shaped waveform the Plasma Display cell can exhibit a number of discrete stable states, each

corresponding to a different slope of the exciting voltage, and therefore to a different intensity. To understand the process, we first review some principles of the Plasma Display cell itself.

The cell, in the "on" state, fires once each half cycle, when the exciting voltage and the wall voltage equal the firing voltage. In equilibrium the charge transferred during each discharge is the same, and the exciting voltage at the times of successive discharges are equal in magnitude, but opposite in polarity. This voltage is called the recurrent volt . Any small change in wall charge causes a corresponding change in recurrent voltage, which, because of the relation between voltage slope and total charge transferred to the walls, reduces the original perturbation. This restoring effect is generally precise and rapid. A recently developed theory (see Section 9.3), however, indicates that with appropriate waveforms, reduction of the perturbation can be made arbitrarily slow, that it can also be made to grow exponentially, and that at the limit of stability, it can persist indefinitely. The deliberate introduction of a perturbation in this case will thus change the state and the intensity.

The three photographs* of Fig. 9.1 illustrate these ideas. The upper traces in each photograph correspond to the exciting waveform itself.

---

*Figures 9.1 through 9.7 are taken from a CSL report now in preparation by W. D. Petty, "Control of Luminous Intensity in the Plasma Display Panel."

The lower trace represents the intensity of the Plasma Cell discharge as recorded on an oscilloscope from a photocell pickup. These traces show that for a duration of 100 μsec the intensity of the cell remains essentially constant. Each trace represents a different intensity level. The photographs do not, however, indicate whether these states are stable or whether they will drift with time. Figure 9.2 shows that the state is unstable, and that after a sufficiently long time the cell will simply turn itself off.

Although a quasi-stable state of this kind would be useful in a television-like display in which a new picture was generated in each basic interval, only stable states are acceptable in the kind of information display we envision. The theory, however, indicates that regions of instability can alternate with regions of stability. Small perturbations from any stable state will then damp out exponentially, and any state once reached will persist indefinitely until it is changed by new information.

The procedure is illustrated in Fig. 9.3 which shows the alternation of stable and unstable sections in the exciting voltage wave. This shape follows from the theory which shows that sections with a positive second derivative are unstable while sections with a negative second derivative, if it is not too large, will be stable.

This wave can be generated by driving an appropriate wave shaping network with a square wave. This process together with a representative circuit is shown in Fig. 9.4.

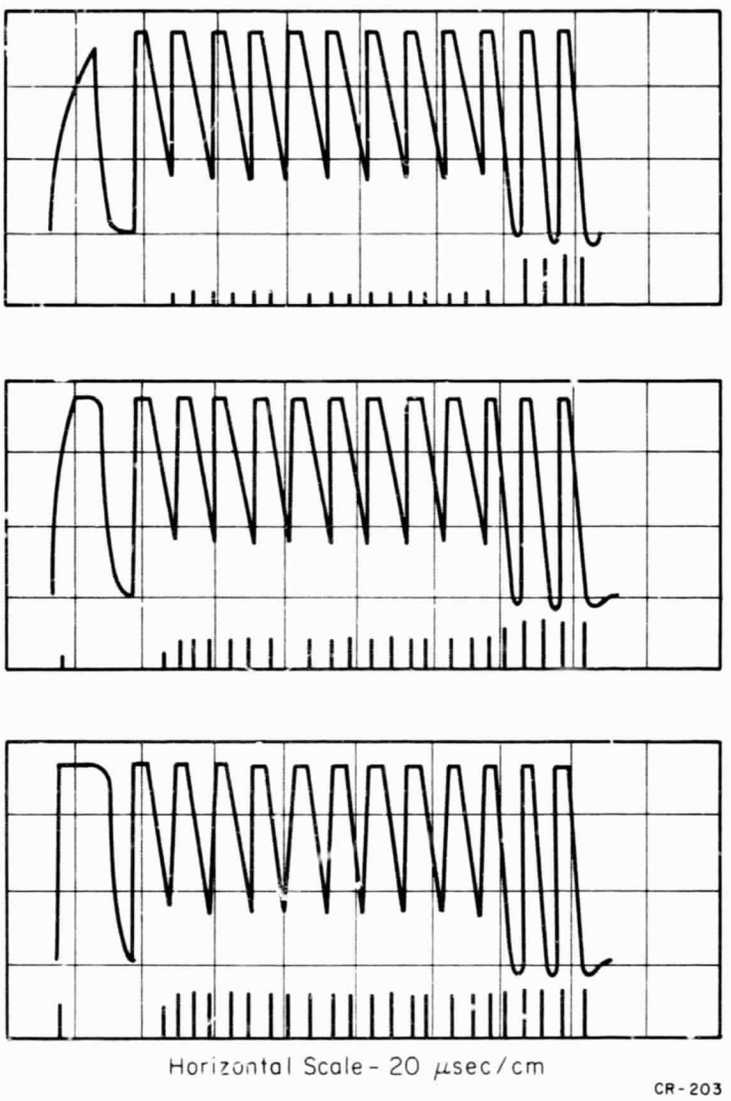


Fig. 9.1. Intensity of cell with modified sustaining signal waveform. Horizontal scale -- 20μs/cm.

(OVERLEAF BLANK)

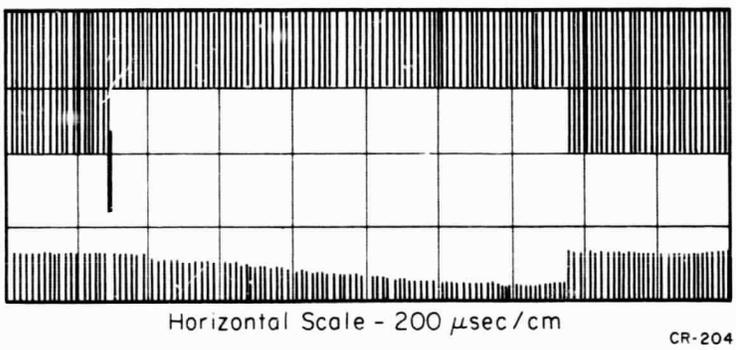


Fig. 9.2. Response of cell to extended duty period. Horizontal scale -- 200μs/cm.

(OVERLEAF BLANK)

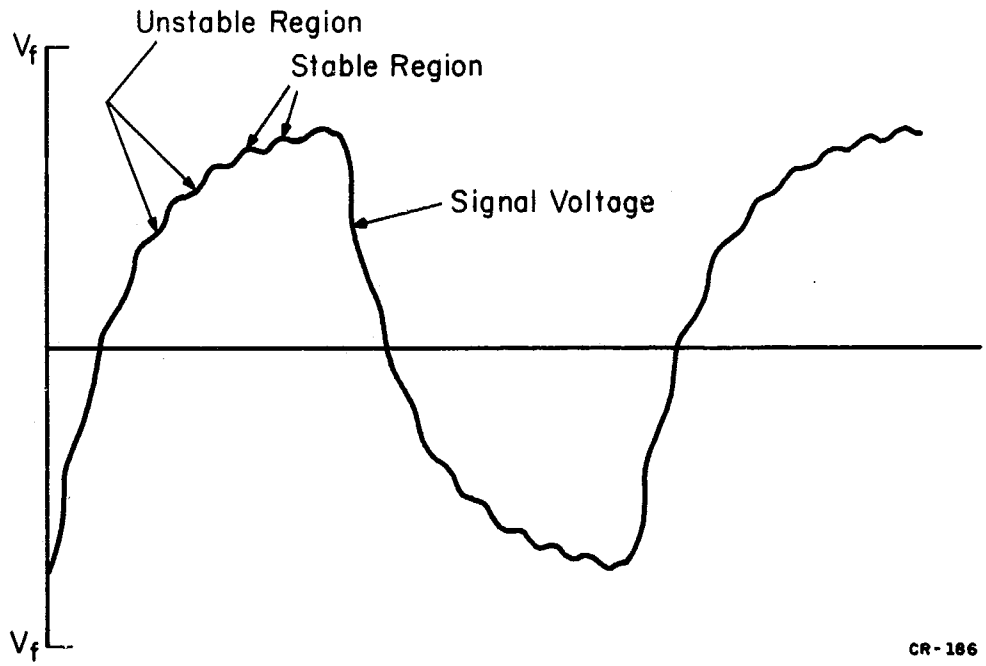


Fig. 9.3. Rippled exciting signal for multiple states.

(OVERLEAF BLANK)

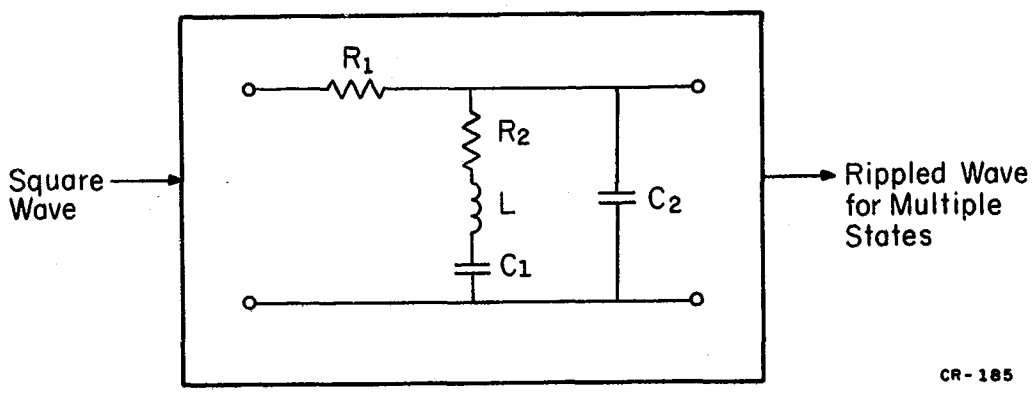


Fig. 9.4. Wave shaping network for multiple intensities for the plasma display cell.

(OVERLEAF BLANK)

9.3. Stability Considerations

The fundamental relation in a discussion of the stability of firing times in the Plasma Display cell is

$$V_{ri} = V_f - V_{wi}, \tag{1}$$

where V_{wi} and V_{ri} are the magnitudes of the wall voltage and the recurrent voltage (the exciting voltage V at the beginning of the i^{th} discharge), and V_f is the firing voltage. The flow of charges to the walls during the i^{th} discharge changes the wall voltage by an amount $2v_{ti}$ to V_{wi+1}, and the next discharge in the sequence is ignited when the exciting voltage reaches the next appropriate recurrent voltage:

$$V_{ri+1} = V_f - V_{wi+1}. \tag{2}$$

Figure 9.5 shows the relations among these voltages for an arbitrary exciting signal. We note, in particular, that since we are concerned with the magnitudes of these voltages, V_{ri} and V_{wi} are always positive, and we, therefore, do not distinguish between the positive and negative half cycles.

Two successive values of wall voltage are related by the expression

$$V_{wi+1} = -V_{wi} + 2V_{ti}, \tag{3}$$

and substituting from Eq.(1), we obtain the relation between two successive values of recurrent voltage

$$V_{ri+1} = 2V_f - V_{ri} - 2V_{ti}. \tag{4}$$

In terms of the difference δ_i between V_{ri} and the equilibrium value V_{r0}, and the difference δ_{i+1} between V_{ri+1} and V_{r0} this relation becomes

$$V_{ro} + \delta_{i+1} = 2V_f - (V_{r0} + \delta_i) - 2V_{ti}, \tag{5}$$

$$\delta_{i+1} = 2(V_f - V_{r0}) - \delta_i - 2V_{ti}. \tag{6}$$

If this process is unstable, δ_i will increase as i increases. On the other hand, if it is stable δ_i approaches zero, and V_{ri} and V_{wi} approach their equilibrium values:

$$\begin{aligned} V_{ri} &= V_{ri+1} = V_{r0}, \\ V_{wi} &= V_{wi+1} = V_{w0}, \end{aligned} \tag{7}$$

The change in wall voltage after each discharge, in equilibrium becomes equal to twice the magnitude of the wall voltage,

$$2V_{t0} = 2V_{w0}, \tag{8}$$

and Eq.(4) reduces to Eq.(1) which we can now write in the form

$$V_{r0} = V_f - V_{w0}. \tag{9}$$

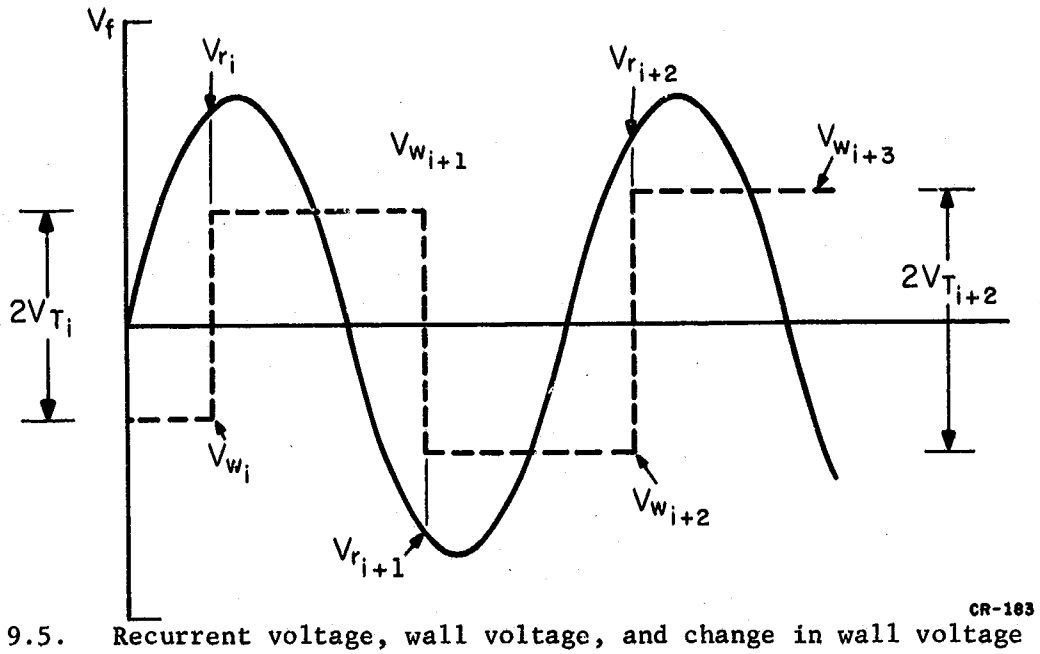


Fig. 9.5. Recurrent voltage, wall voltage, and change in wall voltage during approach to equilibrium.

(OVERLEAF BLANK)

The question of stability rests on the relation between the form of the exciting signal, V, and the change in wall voltage, $2V_T$. This change, in turn, depends on the intensity of the discharge. In general the intensity increases with over voltage. Therefore, since the discharge requires time to develop, the more intense discharge occurs when the slope of the exciting waveform at the time of firing is larger.

We assume now that the change in wall voltage can be expressed as a function of slope. If at equilibrium

$$2V_{T0} = f(\dot{V})\Big|_V = V_{r0}, \tag{10}$$

then more generally

$$2V_{Ti} = f(\dot{V})\Big|_V = V_{ri} = V_{r0}+\delta_i, \tag{11}$$

or, to first order

$$2V_{Ti} = 2V_{T0}+(df/dV)\,\delta_i. \tag{12}$$

From Eq.(7) and (11) we obtain the relation

$$\delta_{i+1} = 2(V_f-V_{ro})-2V_{F0}-(df/dV)\,\delta_i, \tag{13}$$

and substituting from Eq.(9) we finally obtain

$$\delta_{i+1} = -(1+\gamma)\delta_i, \tag{14}$$

where we have let $\gamma = df/dV$.

For values of γ in the range

$$-2<\gamma<0, \tag{15}$$

a perturbation from equilibrium decreases to zero in time. For γ beyond this range, δ grows and the system is unstable.

To separate the influence of the charge-slope relation from that of the waveform itself we replace the term df/dV by its equivalent,

$$df/dV = (df/d\dot{V})[(d^2V/dt^2)/(dV/dt)]. \tag{16}$$

The first factor is a parameter of the cell that can be measured. The second term is the ratio of the time derivative of the slope to the slope itself.

It is instructive to consider the two cases at the limit of stability. When $\gamma=0$, $\gamma_{i+1}=-\gamma$. A perturbation δ then alternates on successive half cycles and persists indefinitely. This condition could arise if the charge transferred were independent of the slope ($df/d\dot{V}=0$), or if the slope did not change in the range that includes both the initial equilibrium voltages and the perturbed firing voltages. This second case is illustrated in Fig. 9.6. When $\gamma=-2$, then there obtains $\gamma_i=\gamma_{i+1}$, and a perturbation δ once introduced, persists indefinitely. The new firing times represent an equilibrium state just as the original firing times do. This case is illustrated in Fig. 9.7. To obtain an acceptable form for

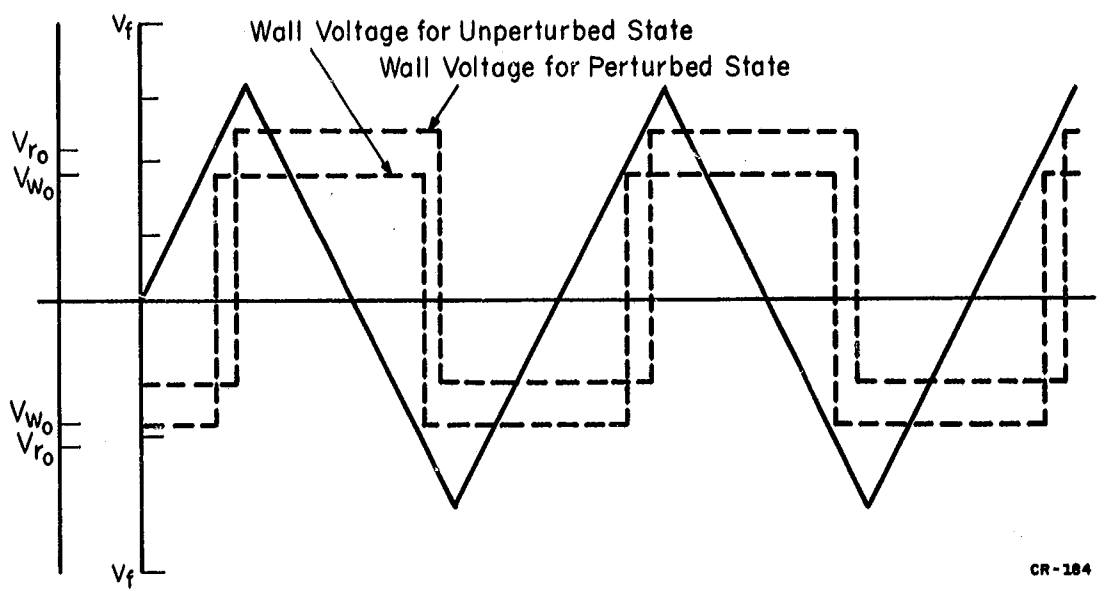


Fig. 9.6. Wall voltage for two stable states with $\gamma = 0$.

(OVERLEAF BLANK)

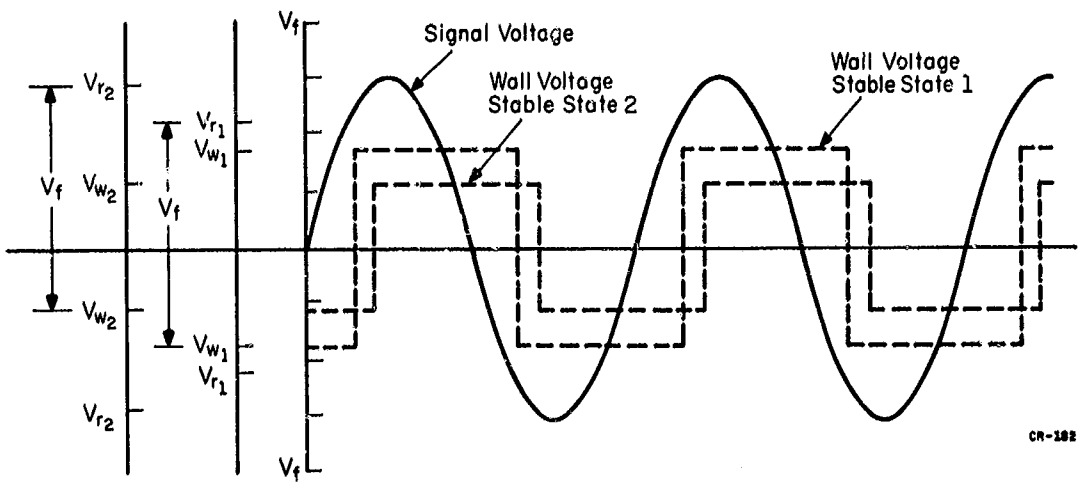


Fig. 9.7. Wall voltage conditions for two stable states with $\gamma = -2$.

(OVERLEAF BLANK)

this voltage wave we must find a solution to the equation

$$(df/d\dot{V})[(d^2V/dt^2)/(dV/dt)] = -2. \qquad (17)$$

If the charge transferred in a discharge were strictly proportional to the slope $df/d\dot{V}$, then a solution would be

$$V = C(1-e^{-2t/k}). \qquad (18)$$

If, as is more typical, $df/d\dot{V}$ is not constant, but varies smoothly with $\dot{V}$, Eq.(18) is still an approximate solution to Eq.(17) over small ranges of $\dot{V}$, provided that k is replaced by the local value of $df/d\dot{V}$ for each range.

A third special case occurs when $\gamma=-1$. A perturbation for this case will be cancelled in the next discharge in the sequence. For values of $\gamma=-1$, therefore, the approach to equilibrium of a cell after change of state will be rapid, a fact frequently observed in work with the Plasma Display cell.

A practical implication of the theory is that through control of the exciting voltage waveform, regions of stability could be alternated with regions of instability. This would lead to a multiplicity of stable states and a corresponding number of intensities.

9.4. A Transformer-Coupled Circuit for Sustaining, Addressing, and Decoding

Driving a Plasma Display Panel requires a voltage of alternating polarity, and of sufficient amplitude, to maintain cells in the lighted state. The amplitude of the current pulses that occur when the panel is in operation depends primarily upon the number of cells actually lighted. If the source driving the panel electrodes is not of sufficiently low impedance, the resulting poor regulation may make satisfactory operation of the panel impossible.

To "turn-off" or "turn-on" individual cells, it is necessary to have unique voltages on a single "x" and a single "y" electrode. The ideal generator for driving Plasma Display Panels is a voltage source, and the least compromise with performance can probably be obtained by having individual voltage sources for each electrode on the panel. However, until semiconductor technology can provide the required devices at low cost, designers will probably turn to some method of signal mixing for addressing display panels. Any of the isolation elements such as resistors or capacitors may be used, but many signal-mixing systems either require excessive power, are too costly, or present too high an impedance to the panel electrodes.

One method for driving panels that is being evaluated at CSL involves the use of multiple-secondary transformers. Given ideal transformers, this approach would provide performance equivalent to a system

with individual voltage generators on each electrode. However, the impedance associated with the leakage inductance of the present transformers is significant. The transformers that have been used so far have employed printed circuit secondary windings. This method of construction would seem to offer good uniformity at a reasonable cost. The transformers, as constructed, are adequate for a four-inch panel if all of the cells associated with a given transformer are not lighted simultaneously. This problem is being circumvented by simply firing groups of cells at slightly different times.

Figure 9.8 is a diagram for this method of operating a Plasma Display. The panel has 256 electrodes in each direction. There are 16 transformers, each with 16 secondaries associated with the "x" electrodes. In addition, corresponding secondaries on each transformer are connected to a common transistor driver so that a total of 16 transistor drivers are required. The circuitry for the "y" electrodes is identical to that of the "x" electrodes.

The circuitry illustrated is capable of lighting or erasing any single cell in the panel and of sustaining the lighted cells. It is also possible to modify 16 cells in two address cycles. This is accomplished by first erasing all 16 cells and then lighting whichever of the 16 cells are desired.

Address decoding is somewhat simplified for this method of operation since only 64 decoder outputs are required, compared to a maximum of 512.

The sustaining voltage signals are illustrated in Fig. 9.9. Since the transformers are not driven during the sustaining operation, they are not required to accomodate large currents, but only the charging currents for the line capacity and the conduction current of a maximum of 16 cells. The transistor drivers must provide any of three voltage levels; e.g., 0, 100, or 200 V.

Figure 9.10a indicates how this technique might be employed to selectively fire a single cell in a panel with 64 cells. Cells in a Plasma Panel fire only once if successive sustaining signals are of the same polarity. The selected cell is x-3, y-3; only transformers x-0 and y-0 are driven; and only the selected cell has 200 V across it with the "x" electrode more positive than the "y" electrode. If the last firing was in a y>x direction, and the signal characteristics are suitable for erasing, then only the selected cell fires and is erased. A similar illustration for lighting cells is shown in Fig. 9.10b. In writing or erasing, no cells have disturbing voltage greater than half the normal sustaining amplitude in a direction that would most easily cause a firing. The transformers are required to provide only unipolar output signals.

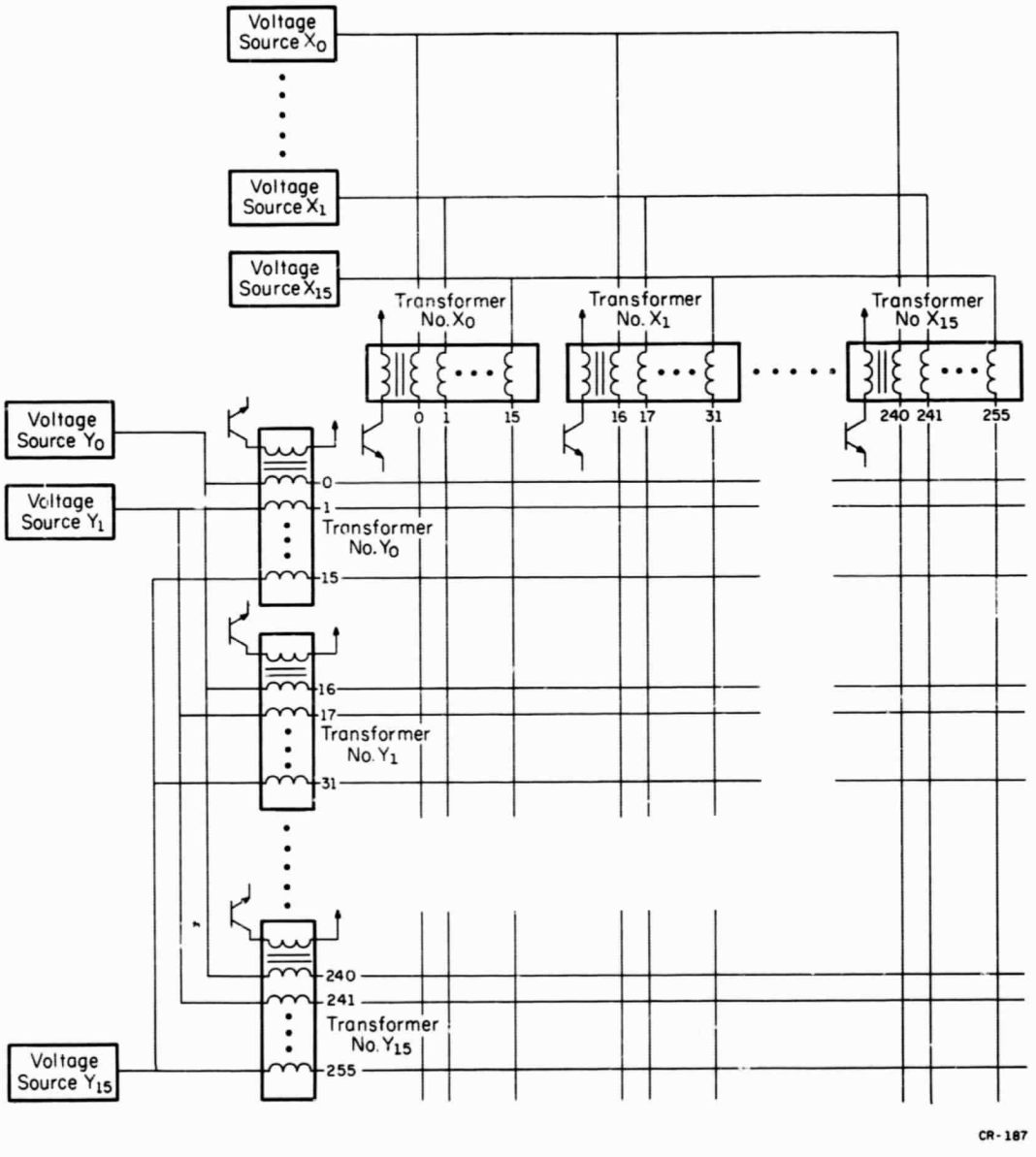


Fig. 9.8. Diagram for driving a 256X256 panel.

(OVERLEAF BLANK)

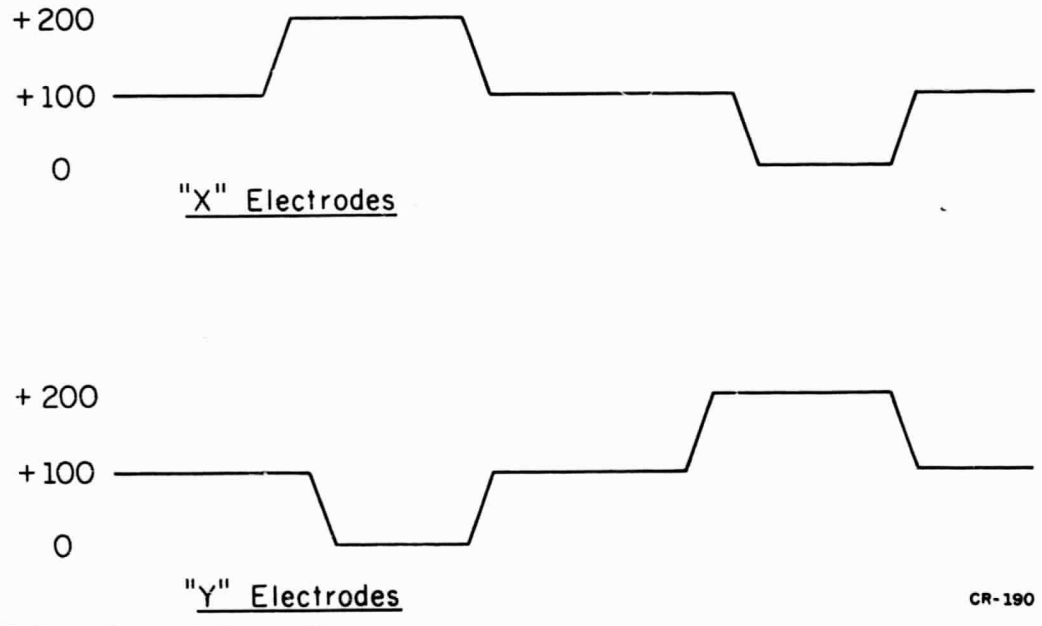


Fig. 9.9. Sustaining voltages.

(OVERLEAF BLANK)

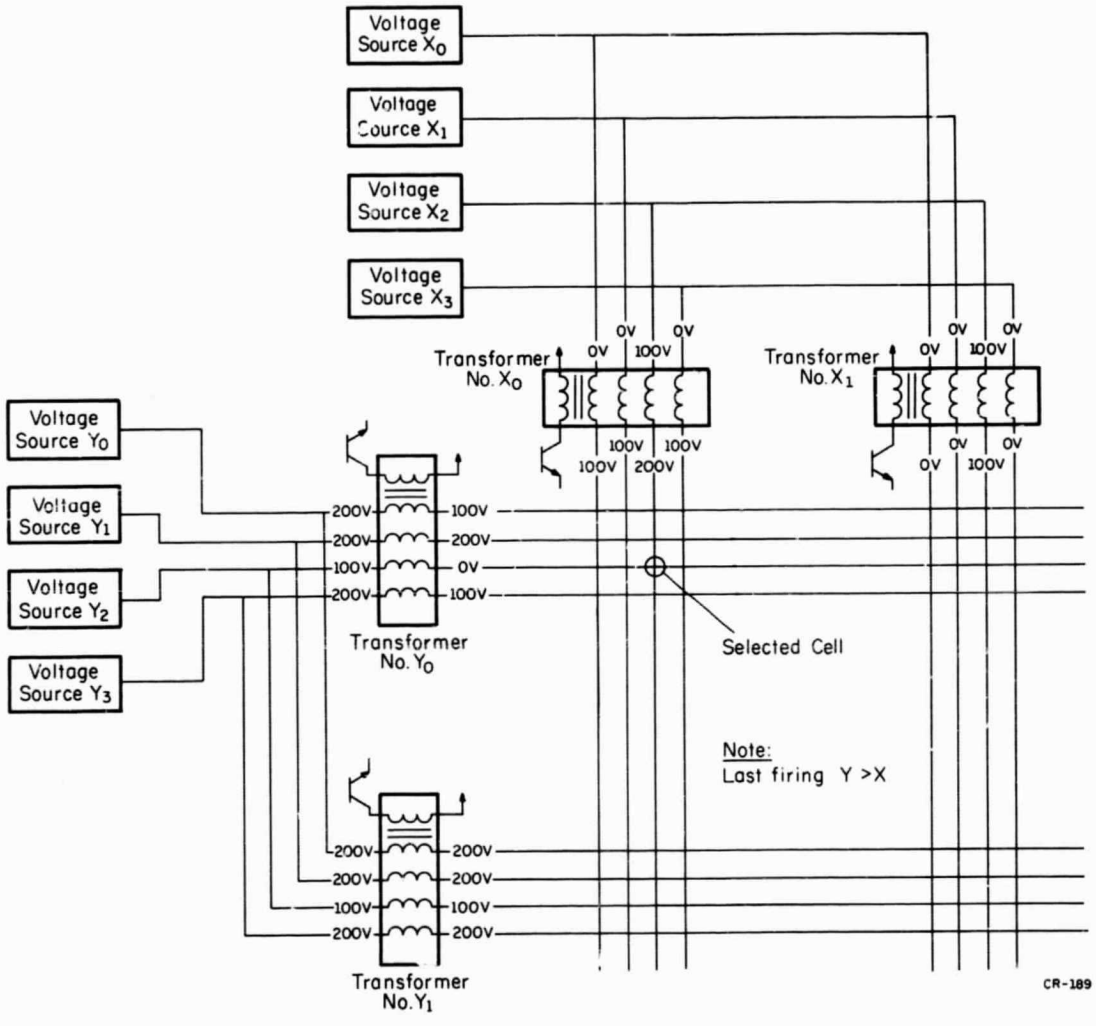


Fig. 9.10a. Voltage addition for erasing a single cell.

(OVERLEAF BLANK)

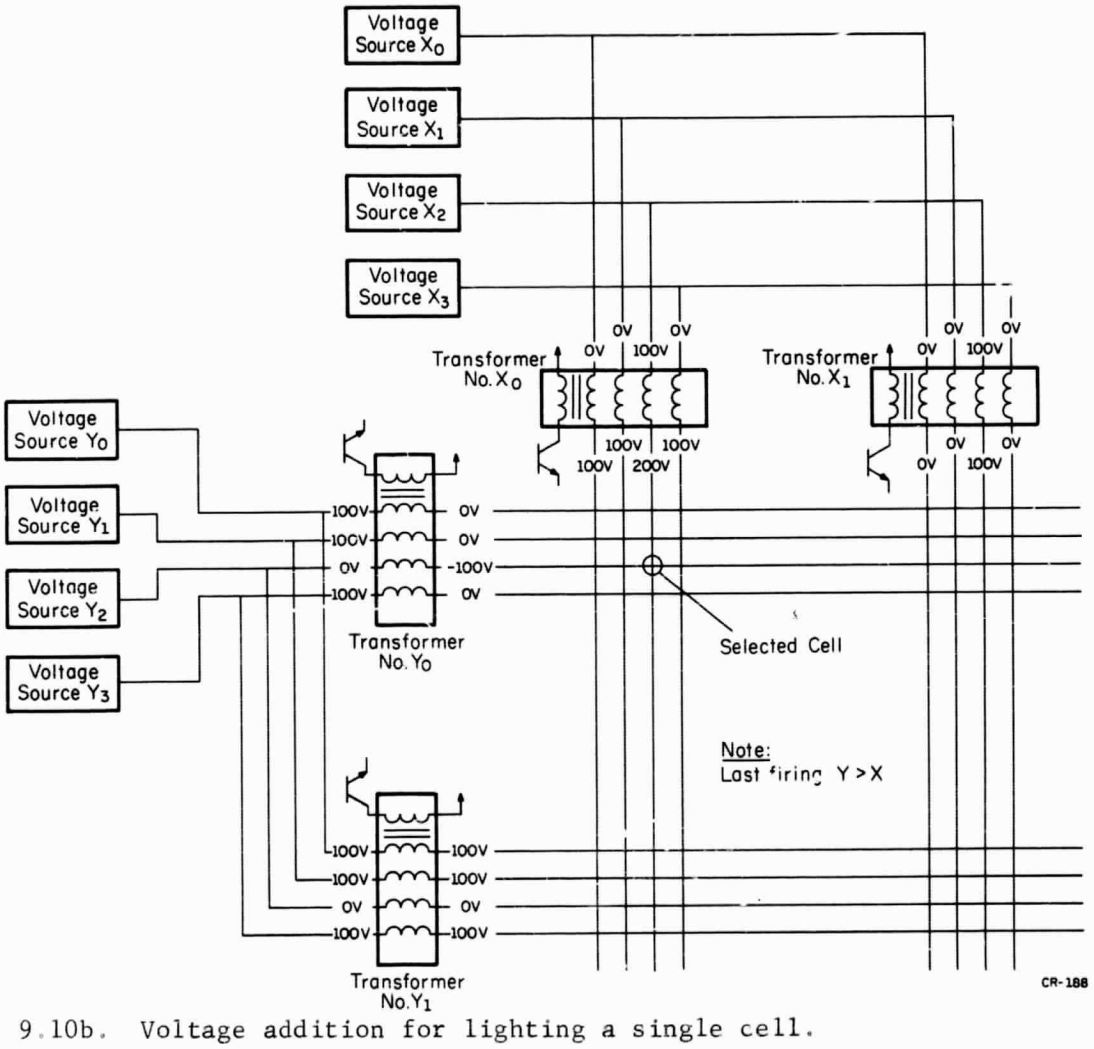


Fig. 9.10b. Voltage addition for lighting a single cell.

(OVERLEAF BLANK)

J. Voss
L. Kadanoff
J. Bouknight
M. Nicholas
W. Brown
R. Lewis*
M. Plunk*
R. Jones*
D. Hammerton*
D. Erbes*

10.1. Project Prospectus

Since this is the first time work by this group has been included in a semi-annual progress report, the first section will outline the history and objectives of this research.

Prior to the academic year 1967-1968, Professor Voss was conducting a study of a number of towns planned by the Illinois Central Railroad. The purpose of this research was to examine a limited number of hypotheses concerning the factors which determined the sale price of urban parcels. One of the key variables studied was accessibility and the way this influenced land values. In order to study this problem, a considerable amount of land-sale-price data were collected for approximately 27 communities. Many time-sequence maps were prepared by hand and the statistical analyses were done without immediate reference to spatial location. In the fall of 1967, this project was moved into the Coordinated Science Laboratory because it was felt that the computer-display facilities of the Laboratory could greatly enrich and increase the analysis opportunities. As work progressed,

---

†This work was supported by Ford Motor Company.

*Student assistants employed during Summer of 1968.

it became evident that our explorations offered the opportunity of developing not only vastly improved techniques and methods of analyzing urban spatial structure but making some real contributions to the understanding and theory of urban growth and decay. In addition, this early work also indicated the value of examining city growth by individual parcels and the manner in which this could be done. The data used were the amount, location, and date of each transaction. This information was obtained from the Recorder's Office in the several counties. A source of information was used that had heretofore been ignored, and yet the data sharply revealed the dynamics of urban growth.

Based on the experience of studying and graphically analyzing these small towns, it was decided that we should undertake a comprehensive study of one city, Kankakee, Illinois (see Figs. 10.1 and 10.2).

The scope of this project can be divided into two general sections. The first of these emphasized analysis aimed at contributing to our understanding of urban structure and its causes. The second functions as a means to these ends by developing new analysis techniques utilizing the graphic-display capabilities of high-speed computers.

Kankakee was selected because it seemed to be of manageable size (approximately 50,000 persons); it was reasonably close to Champaign-Urbana (about 1 hour and 30 minutes driving time); both its social and economic structure were and have been relatively independent and self contained; it has significant prospects for future growth since it is near Chicago and with the completion of I-57 will be approximately an hour's drive from

Fig. 10.1. State of Illinois. Circle indicates location of Kankakee.
Source: U.S. Department of Interior, Geological Survey.

(OVERLEAF BLANK)

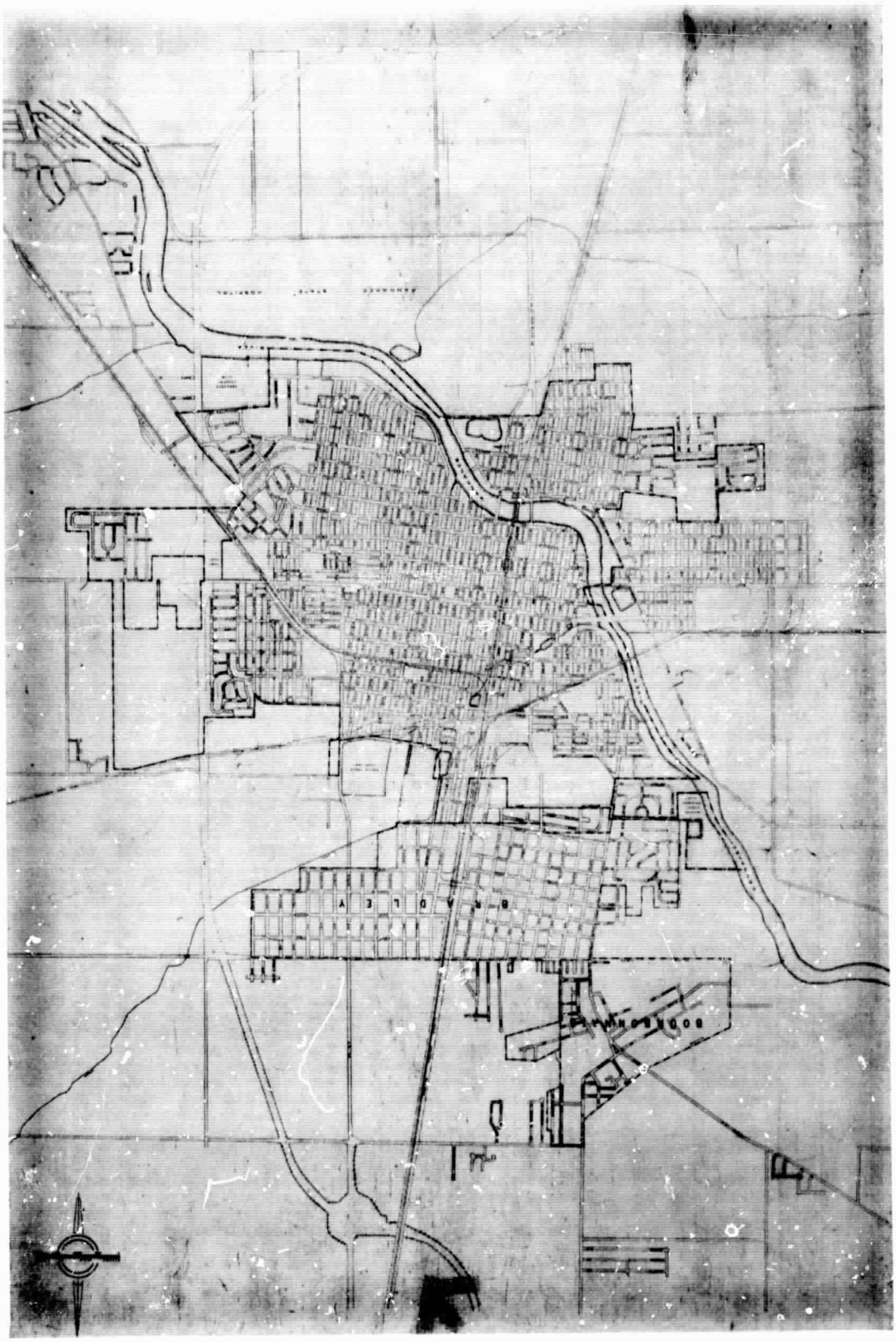


Fig. 10.2. Kankakee, Illinois. Source: Harlan, Bartholomew and Associates, St. Louis, Missouri.

(OVERLEAF BLANK)

downtown Chicago; it offers a wide variety of topographic features ranging from flatlands to hills and bluffs overlooking the Kankakee River; it contains a broad range of urban characteristics including all types of land uses and socio-economic neighborhoods; it provides a complete mix and history of all modes of transportation; its physical layout ranges from the gridiron in the city center to free-flowing curvilinear streets on the periphery; and finally, Kankakee was selected because good, detailed data throughout its entire history were available and could be obtained with a reasonable expenditure of effort.

The analysis portion of this study can be viewed as moving through three stages of development. During the first stage, which we anticipate initiating in November of 1968, we will carry out an in-depth analysis of the information presently being assembled. More specifically, we will integrate land use and structure information with the land value and transaction data. We will try to develop an understanding of how various blocks and subareas of the city changed through time. Simultaneously, we will be viewing this process in the light of the transportation economics, amidst changes that were taking place. The aim of this effort then is to structure the guidelines for the subsequent examinations of Kankakee using existing theories of urban growth and decay.

The second phase of the analysis will involve testing selected growth models using the Kankakee data. The purposes of this portion of the study include identifying those theories that are most efficacious and establishing the general framework for new theoretical developments. In

this stage, we will be examining broad theoretical concepts as well as particularities such as the parameters associated with a specific variable or the different levels of areal aggregation. The work during this phase becomes closely interlocked with the final stage, the production of new conceptions and ideas about urban growth. One of the most exciting aspects of this project is the opportunity to move away from the static models of urban growth and into the modeling of cities as they actually are, namely, dynamic processes of spatial and nonspatial growth and decay. This, then, represents our ultimate objective.

The types of analyses and study that are contemplated will require modifications to our methods of handling these data and interacting with them through the computer-graphics media. Although the Coordinated Science Laboratory nas an extensive computer-graphics facility, the presentation of large quantities of data in a time sequence with meaningful interaction capabilities requires a considerable programming effort. These programs must be developed such that analysis can be implemented and the results directly compared with historical data.

The results of this detailed effort in computer-software development will be to make available a highly flexible system that can be used for analysis of historical land-usage data for essentially any city in the country. The animation facility will be utilized to present the historical data in any time sequence. The capability of presenting various alternatives for future growth will be available with the patterns of future development being determined by various factors related to transportation systems, industrial location, etc.

Presently, information is being collected on the date, type, and sale price of each real estate transaction, from 1855 to the present, in alternate city blocks throughout the metropolitan area. The sources of these data are the offices of the local abstract company and the County Recorder. At the same time, we are assembling detailed data on a number of social and economic variables, such as industrial growth, changes and improvements in the transportation system, school construction and policies, social and demographic changes, and provision of public utilities. Much of the date will be punched onto IBM cards by that date. Upwards of 70,000 pieces of information will be available in machine-readable form with the completion of the key punching.

It is the intent of this project to investigate, in depth, the factors that led to the development of the Kankakee urban area. The identification of the pertinent factors and the evaluation of their relative importance will provide the means for a detailed testing of land-use models. In order to do this, it will be necessary to develop a highly flexible, computer-graphic technique for presenting and analyzing the historical data and the predictions that result from the models. Although the results of this study will be of obvious importance and value to the City of Kankakee, it should be emphasized that the central objective of this research is to produce findings applicable to all types of urban areas.

10.2. Real-Estate Transactions

10.2.1. Base Data

The following data have been collected from the Kankakee Title and Trust Company for each lot in alternate blocks throughout the entire metropolitan area:

a. Date of transaction,

b. Type of transaction,

c. Block number,

d. Lot number within each block (see Fig. 10.3),

e. Subdivision number,

f. Amount of transaction in current dollars,

g. Book number in County Recorder's Office containing transaction,

h. Page number in book containing transaction.

10.2.2. Missing-Price Data

Missing Price Date--From time to time the price of a deed was not recorded at the time the deed was filed. When these events occur, the actual deed is checked in the Recorder's Office and the transaction amount is determined from the affixed tax stamps.

10.2.3. Data Format for Graphic Display and Analysis

After the data from the Kankakee Title and Trust Company have been collected, they are re-arranged and supplemented with additional data before punching onto cards. The initial data cards contain the following information for each transaction:

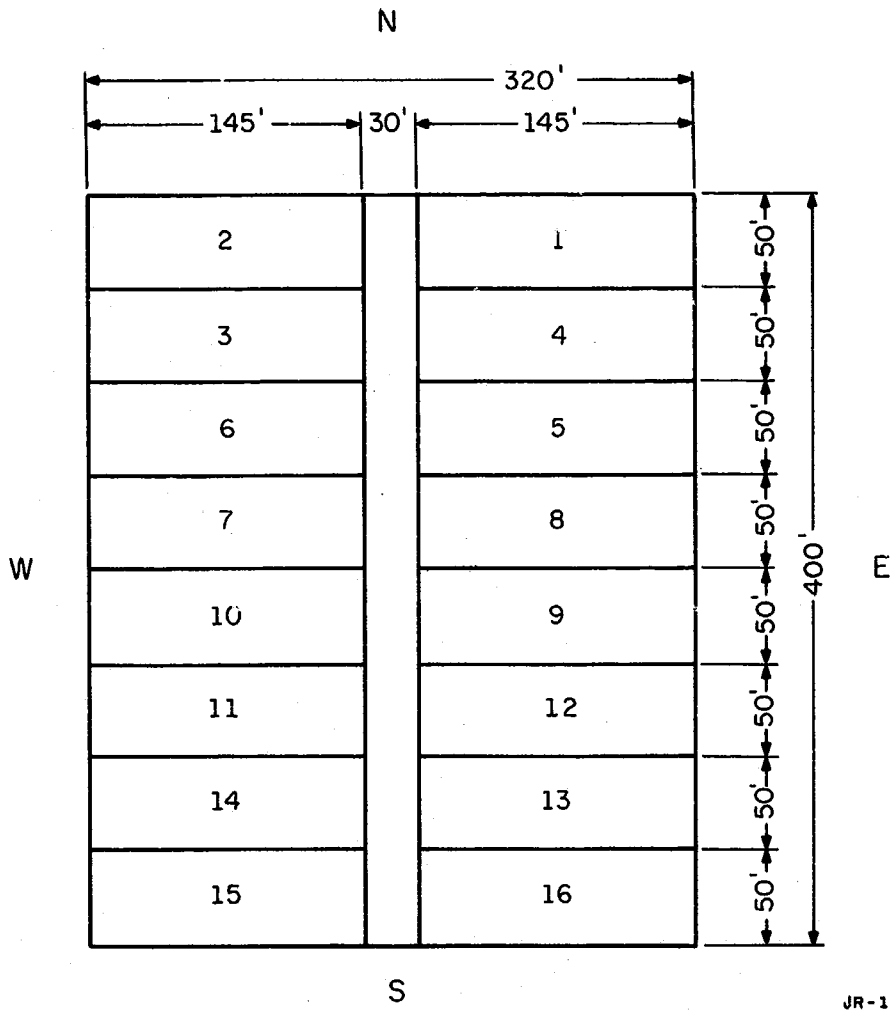


Fig. 10.3. Typical lot arrangement within blocks. Source: Original plot of Kankakee by Hiriam Ketcham, January 14, 1854 as recorded in Kankakee County Recorders Office.

(OVERLEAF BLANK)

a. Subdivision number,

b. Block number,

c. Lot numbers within block in multiple-lot transaction,

d. Number of first lot in transaction,

e. Total amount of block in transaction,

f. Month of transaction,

g. Year of transaction,

h. Transaction location within a block,

i. Type of transaction,

j. Price to nearest dollar,

k. Book number in County Recorder's Office,

l. Page number in Book,

m. Document number (transaction number),

n. Area in square feet of transaction,

o. Street E-W code.

Data cards are punched for each transaction, sorted and listed and additional warranty deed, trust deed, and deed data are collected from the County Recorder's Office. The additional data cards are merged with the original data deck and a matrix is produced which chains all lots in a clock from the first sale to the present time (see Figs. 10.3 & 10.4).

10.3. Land-Use, Structure, and Dwelling-Unit Data

The transaction matrix is used in conjunction with Sanborn Insurance Maps of Kankakee in order to produce additional information

Fig. 10.4. (See facing page.) Chained transactions within a block.

Source: Kankakee Title and Trust Company and County Recorder's Office.

Legend - for Computer Print-Out

Dates - Date at which a transaction was recorded

Amounts - The amount involved in the transaction to the nearest one dollar.

Type of Transaction:

W - Warranty deed

T - Trust deed

D - Deed

M - Mortgage

P - Part of lot included in transactions
(if no P indicated then entire lot sold)

There is only one transaction on each horizontal line.

Lots have been arranged according to the actual plat of the block.

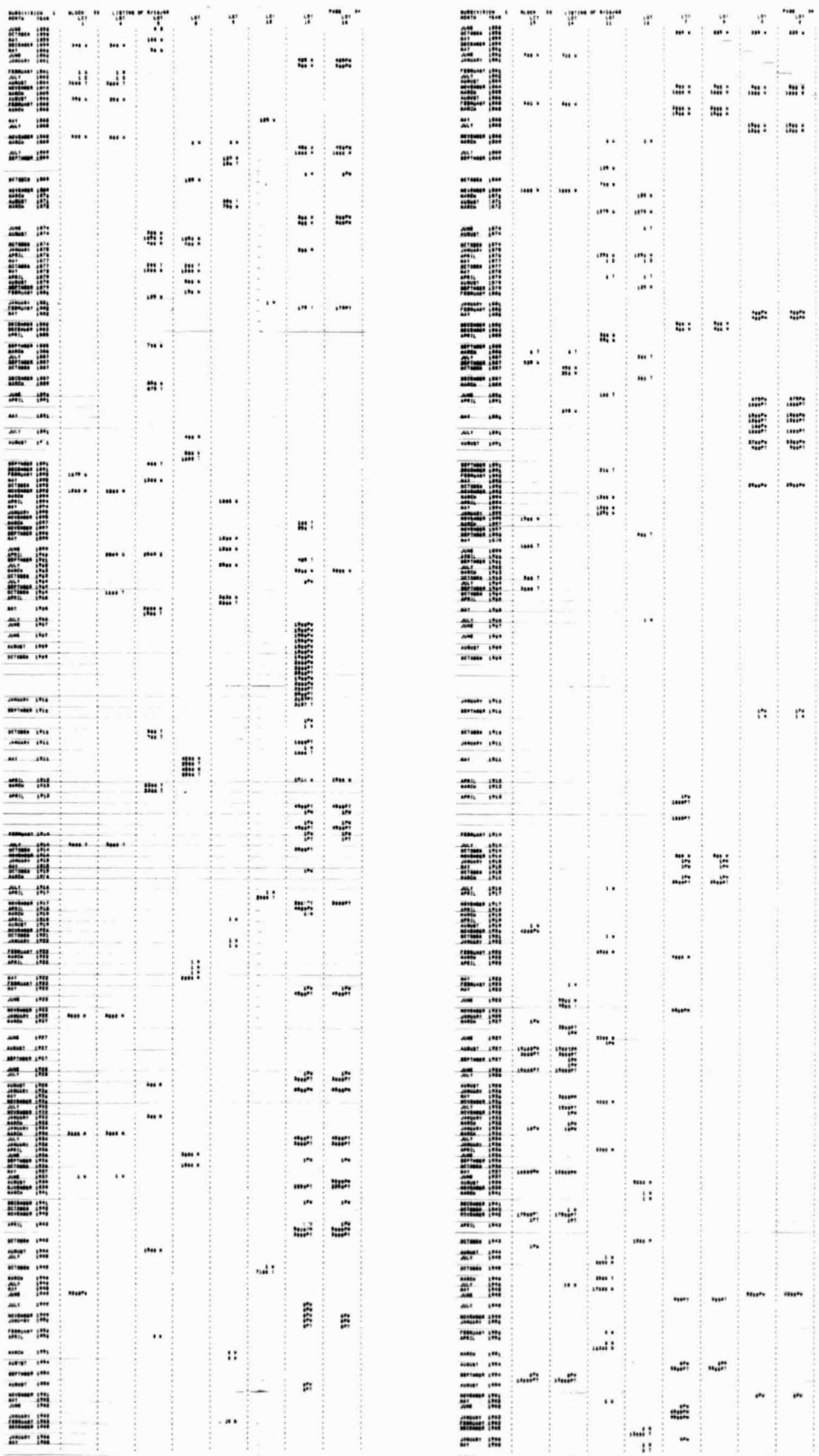


Fig. 10.4. Chained transactions within a block. Source: Kankakee Title and Trust Company and County Recorder's Office.

(OVERLEAF BLANK)

on history of land use and structures for each block. The following tasks are in the initial stages of development.

10.3.1. Land Use

For each half block and block, the changing number of square feet devoted to the following uses; residential, industrial, commercial, and public will be calculated.

10.3.2. Residential Structures

For each half block and block, a running inventory of residential structures will be developed.

10.3.3. Dwelling Units

For each half block and block, a running inventory of the number of dwelling units will be developed.

The above data will be ultimately merged with the original transaction information for each block.

10.4. Computer Mapping

Photographic copies of every recorded subdivision in the Kankakee Metropolitan area have been assembled from the County Recorder's Office. From 1854 to the present, 394 subdivisions have been recorded. These data have been aggregated into ten-year periods and placed on overlays of the metropolitan area. Using an X-Y coordinate system, we have recorded the following coordinates for each subdivision; boundaries,

block centroids, street center lines, scale factors, date of plat and dates of any subsequent changes, and platted public facilities, such as parks, etc. This information will be used to produce a time sequence of the physical growth of Kankakee using the display techniques developed for the Illinois Central Railroad towns. Ultimately, the data described in 10.2 and 10.3 will be displayed and analyzed within this structure.

10.5. Variables Affecting Land Use and Development

Much historical information has been assembled about a number of variables which may have influenced the development of Kankakee generally and, in particular, land values. An effort is being made to develop a running inventory of this information so that it will be possible to produce time sequences of all variables. Information has been collected in varying degrees of completeness from 1854 to the present for the following categories: transportation facilities, sewers, water, parks, schools, industry, churches, social and economic status, commerce, ethnic and racial characteristics, and other miscellaneous facilities.

10.6. Graphical Display of First-Sale-of-Lot Data for Towns Developed by the Illinois Central Railroad

A first step in the application of graphic display to the analysis and reduction of data on first sale of lots for towns laid out by the Illinois Central Railroad in the period 1850 to the present was made during the reporting period.

10.6.1. Purpose

The purpose of the graphic display is to produce time sequence studies of the land sale data in four categories:

a. All lots in a town (see Fig. 10.5),

b. Residential lots in a town (see Fig. 10.6),

c. Non-residential lots in a town (see Fig. 10.6),

d. Corner lots in a town (see Fig. 10.6).

10.6.2. Information Displayed

a. An overlay map of the town with the following features displayed (see Fig. 10.5):

   i. City limits,

   ii. Streets,

   iii. Railroad trackage and right of way,

   iv. Block boundaries,

   v. Graphic aids such as churches, schools, etc.

b. A point of light at the center of each lot appearing at the point in time when the lot was first sold conditional on the type and price of lots being displayed for a given sequence.

c. The display may also be made for lots in one of the available categories on the basis of price of sale with both a minimum and maximum value of price selected.

d. For more detail study of a given town, an inset feature is available where a portion of the display may be selected and blown up (see Fig. 10.6).

e. Four graphs of information may also be obtained (see Fig. 10.7):

   i. Number of lots sold on a given date,

   ii. Percent of lots sold to a given date,

   iii. Number of lots sold at a given price,

   iv. Percent of lots sold up to a given price.

f. Various other features of the system include the making of movies of the sequences and the capturing of individual frames of a sequence on various photographic media.

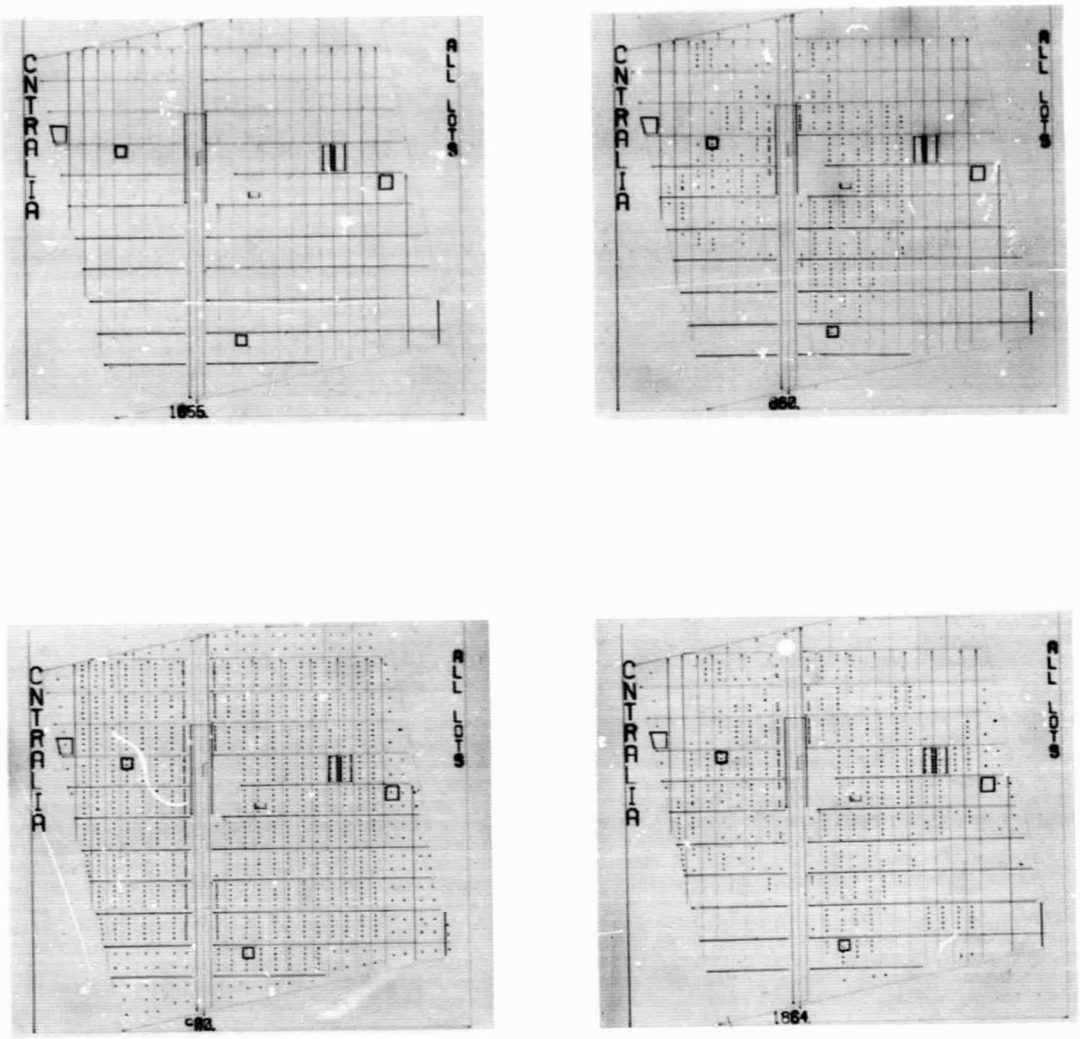


Fig. 10.5. Land sales in Centralia, Illinois. Outlined parcels indicate public facilities and each dot represents the location of one sale. Source: Plat of Centralia by Illinois Central Railroad and Land Sales Records of the Illinois State Archives, Springfield, Illinois.

(OVERLEAF BLANK)

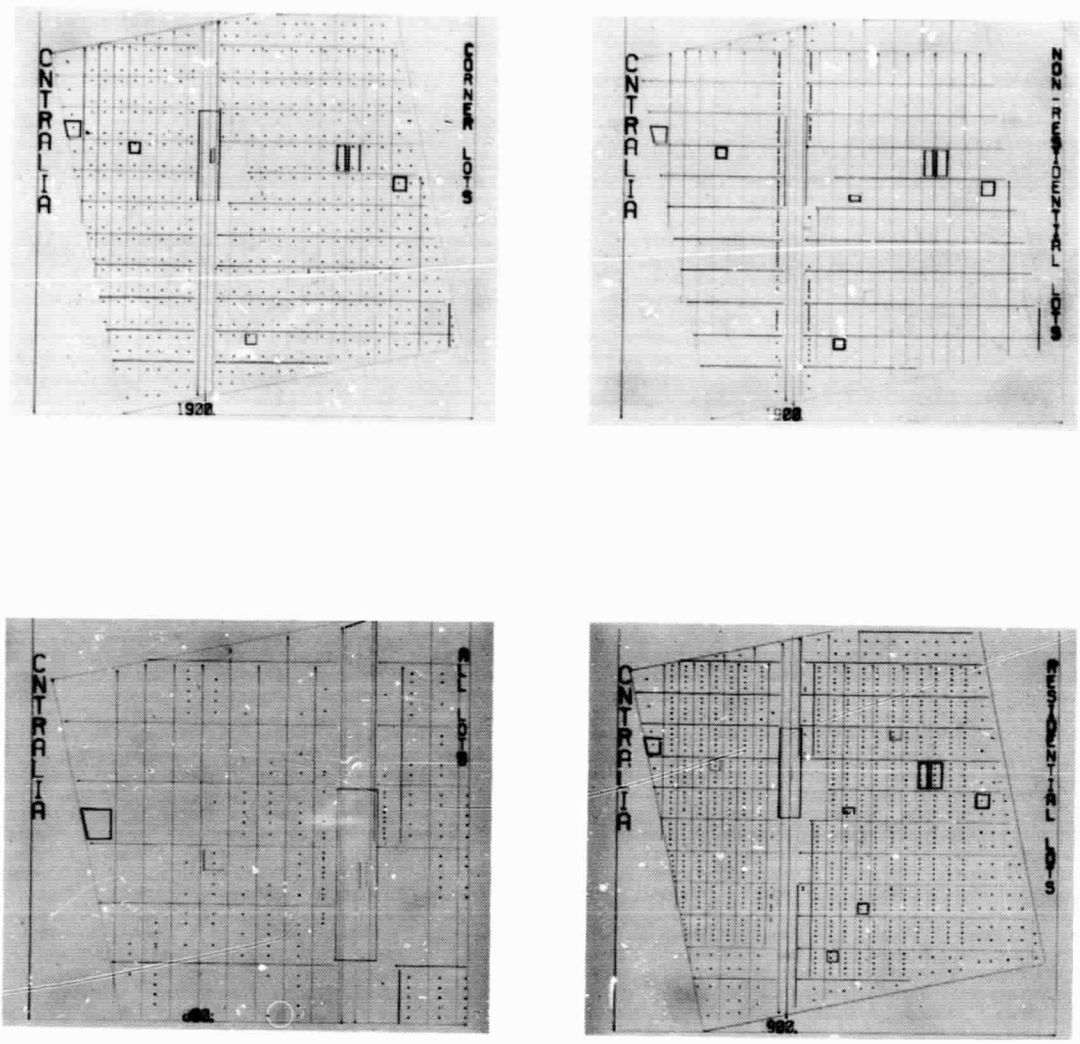


Fig. 10.6. Land sales in Centralia, Illinois showing corner lots, non-residential lots, residential lots and enlarged inset. Outlined parcels indicate public facilities and each dot represents one sale. Source: Plat of Centralia by Illinois Central Railroad and Land Sales Records of the Illinois State Archives, Springfield, Illinois.

(OVERLEAF BLANK)

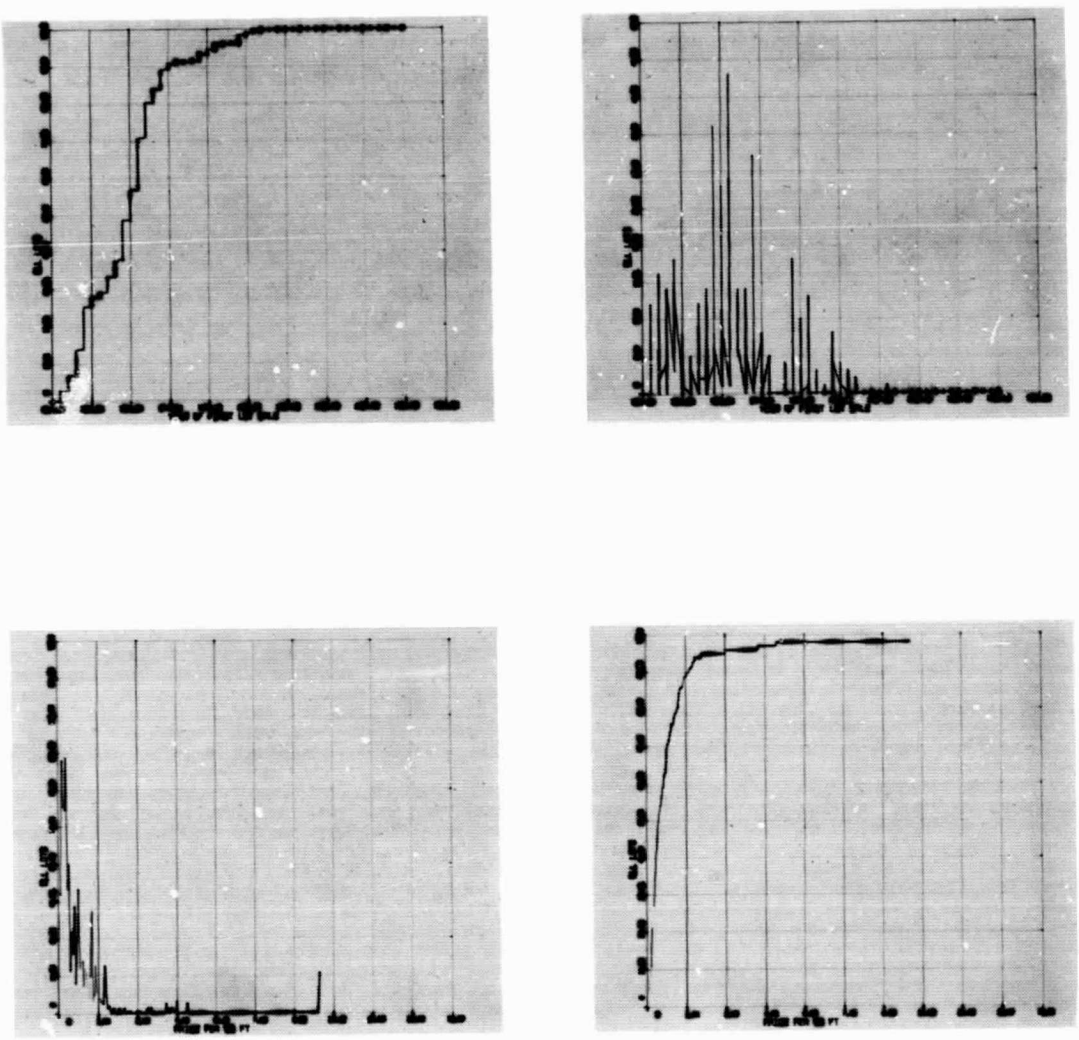


Fig. 10.7. Land sales graphs of Centralia, Illinois in constant dollars. Source: Land Sales Records of the Illinois State Archives, Springfield, Illinois.

(OVERLEAF BLANK)

R. T. Chien
R. B. Ash
D. E. Carroll
D. H. Cooper
A. H. Haddad
E. R. Lyman
F. P. Preparata
S. R. Ray
L. R. Bahl
E. Beach
C. Hartmann
S. J. Hong
R. J. Kubala
G. J. Lipovski
J. K. Moore
S. W. Ng
K. K. Tzeng
T. Woo

11.1. Coding Theory

11.1.1. Syndrome-Transformation Decoding

On decoding a cyclic code, the syndrome polynomial s(x), which is the remainder when the received polynomial r(x) is divided by the generator polynomial g(x), may first be calculated at the receiver. For an (n,k) q-ary cyclic code, it is known that there are q^{n-k} distinct cosets of the group forming the code and q^{n-k} syndrome polynomials, each corresponding to a particular coset. Furthermore, correctable error patterns are coset leaders in distinct cosets. When s(x) is nonzero, then errors have been introduced by channel noises and a decoding procedure is required to find out the error locations and error magnitudes from this syndrome. However, only those syndromes which correspond to guaranteed correctable error patterns can errors be corrected and codewords be obtained at the output of a decoder. For the rest of the syndromes which do not correspond to correctable error patterns, the errors can only be detected. Therefore, for a t-error-correcting cyclic code, the number of such uncorrectable error syndromes is

$$q^{n-k} - \sum_{i=0}^{t} \binom{n}{i},$$

†This work was supported by National Science Foundation under Grant GK-2339 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

which can be quite a large number for large n and t. Inability to use these syndromes to correct more than t errors has been considered a drawback of all algebraic decoding algorithms.

In the last few months, a method for correcting some of the "uncorrectable" error patterns has been investigated. It can be shown that an uncorrectable error pattern will become correctable if a certain transformation exists for this pattern. The whole scheme has been referred to as syndrome-transformation decoding. We illustrate the concept of syndrome-transformation decoding by showing its application to cyclic codes for solid-burst-error correction.

A binary, solid-burst error of length b is defined as an error vector with b adjacent nonzero components. The corresponding error polynomial considered as an element in the algebra of polynomials modulo $x^n - 1$ can be written as $e(x) = x^k(1+x+x^2+\ldots+x^{b-2}+x^{b-1})$ where $0 \leq k < n$ and $1 \leq b < n$. It is believed that codes which correct solid-burst errors may find important application in synchronous detection schemes when the carrier is suppressed. [1] Previous work on binary solid-burst-error-correcting codes has been done by Wyner [2] and Schilinger. [1] The codes proposed by Wyner to correct single solid bursts of length up to b are generated by $g(x) = (x^c+1)p(x)$ where p(x) is an irreducible polynomial of period e, e does not divide c and $e \geq b+1$, $c \geq b+1$. The number of parity-check digits is therefore greater than b+1. On the other hand, Schilinger has synthesized a class of binary single solid-burst-error-correcting codes whose generator polynomial is $g(x) = x^m p(x)p(1/x)$ where p(x) is a primitive polynomial of degree m.

The codes are shown to be able to correct single solid bursts of length up to $b = \frac{1}{2}(n-1) = 2^{m-1}$ within a code length $n = 2^m-1$ and use only 2m parity-check digits. However, it is observed that the problem of solid burst-error-correcting is actually equivalent to finding the starting and ending locations of the burst and hence can be very easily treated as a random error-correcting problem and the class of codes synthesized by Schilinger is indeed the class of double-error-correcting (t=2 only) cyclic codes already discovered by Melas. [3]

A very general case of codes that correct single solid burst of any length and an extension to multiple solid-burst-error correction can thus be constructed. It has been proved that if g(x), not divisible by x+1, generates a cyclic, binary, double-error-correcting code, then the code generated by (x+1)g(x) can correct single solid bursts of any length. Extension to multiple solid-burst-error correction is then immediate and it can be very easily shown that if g(x), not divisible by x+1, is a t-error-correcting code, then (x+1)g(x) generates a code which can correct up to $s = \lceil t/2 \rceil$ solid bursts of any length within a codeword, e.g., $s = \lceil (d-1)/4 \rceil$ or $d \geq 4s+1$, when the minimum distance of the code (x+1)g(x) is d.

A simple decoding procedure has also been devised. The details of the syndrome-transformation decoding scheme have been summarized in a short paper.

K. K. Tzeng
R. T. Chien

11.1.2. Multiple-burst-correcting codes

A very realistic and practical model for many transmission [4] and storage [5] channels used today is one where the error patterns occur as a number of short bursts within one code word. A number of methods for correcting multiple bursts have been suggested, the Reed-Solomon [6] codes being the most useful. A new class of cyclic codes for correcting multiple bursts has been constructed. Details are being written up and will be available in a technical report shortly.

R. T. Chien
L. R. Bahl

11.1.3. Distance Structure of Nonprimitive BCH Codes

An investigation is being made in determining the distance structure of nonprimitive BCH codes. Recent computer analysis indicated a strong possibility that nonprimitive BCH codes possess much higher distances than the value predicted by the BCH bound.

C. Hartmann
R. T. Chien

11.1.4. A Class of Optimum Double-Error-Correcting Nonlinear Codes

This investigation has been motivated by the recent discovery, made by Nordstrom and Robinson, [7] of a nonlinear (15,8) binary code which has the largest number of code words compatible with its length and minimum distance (d=5). After the (15,8) code was presented, two main questions, very intimately related, were posed: 1) What is the mechanism which generates the distance, so far heuristically assessed? 2) Is the NR code a member of

a general class of double-error-correcting codes? Both of these questions have received a complete answer.

The first step was the recognition that the (15,8) code could be described in terms of polynomial codes, i.e., cyclic codes, thereby bringing to avail a considerable body of knowledge. Specifically the (15,8) code was representable as $[m(x)+q(x),i, m(x)+(m(x)+i)u(x)+q(x)f(x)]$, where $m(x)$ is a member of the (7,4) cyclic Hamming code $\{m(x)\}$, $q(x)$ is a coset of $\{m(x)\}$, i is a binary parameter, $u(x)=(x^7+1)/(x+1)$, $f(x)$ is the generator of the dual code of $\{m(x)\}$. This representation indicated that the (15,8) code resulted from the union of a linear code (kernel), that is, $[m(x),i,m(x)+(m(x)+i)u(x)]$, and of seven of its cosets. With the help of the same description it was possible to give a rigorous proof of the distance generating mechanism. [8]

The second question has been answered by recognizing that the kernel has actually the form $[m(x),i,m(x)+(m(x)+i)u(x)+s(x)]$ where $s(x) \in \{s(x)\}$, $\{s(x)\}$ being the distance 6 BCH code ($\{s(x)\} \equiv 0$ for the (15,8) code, which prevented the generalization for some time). It has been found that these codes, of length (2^n-1), exist only for even n.

The interesting features of these codes can be summarized as follows: 1) They contain twice as many code words as the double-error-correcting BCH codes of the same length; frankly, this would be a negligible gain were it not that 2) they have the largest number of code words possible for given length and distance, i.e., are optimal; 3) decoding is based on the calculation of syndrome-like quantities and its complexity is comparable to the one of the corresponding BCH codes; 4) the codes are systematic and

encoding can be accomplished very simply by shift registers in as many time units as are required by the serial transmission of the information digits.

Further generalization in the direction of nonlinear multiple-error-correcting codes, which might be superior to the corresponding linear codes, is the object of continuing investigation.

F. P. Preparata

11.2. Arithmetic Codes

Arithmetic codes have the form AN, where A is a fixed integer generator, and N is the source word ranging from zero to some maximum number of codewords M. Arithmetic codes are designed to correct the mistakes in any computing machine's arithmetic operations. The arithmetic weight of a codeword is defined as the least number of nonzero terms necessary to express the codeword in a modified binary form, where both positive and negative ones may appear. An error is t-fold if its arithmetic weight is t, and the code can correct errors up to t if the minimum distance of the code dm is larger than 2t. Distance between the codewords AN_1 and AN_2 is the weight of $|AN_1-AN_2|$, and the minimum distance is easily seen to be same as the weight of the least-weight codeword.

Following the discovery of the generator of the form, $A=(2^e-1)/B$ (B=number of codewords, e=code length=exponent of 2 modulo B). A considerable amount of study was done in this direction by the authors. [9,10,11] Continuing on the same approach as in the previous Progress Report (April 1968), we narrowed down the search area for B to be the area between the saturation product of given n primes and the point, $B=\Pi_i p_i$, for $1 \leq i \leq n$. The following is a short table of codes we found.

Table 11.1. List of codes discovered.

Code length	B	Minimum distance	Efficiency (> 1/3)
10	11	4	0.400
12	39	4	0.500
14	43	4	0.429
15	151	4	0.533
18	133	5	0.444
18	657	4	0.556
20	123	6	0.350
20	451	5	0.450
20	3813	4	0.600
21	337	5	0.429
21	2359	4	0.571
22	267	6	0.409
22	15709	4	0.636
24	663	6	0.417
24	1989	5	0.458
24	46995	4	0.667
25	601	7	0.400
25	1801	5	0.440
25	55831	4	0.640
26	2731	4	0.462
28	1247	7	0.393
28	1695	6	0.393
28	24295	5	0.536
28	215265	4	0.643
29	2089	6	0.414
29	486737	4	0.655
30	1661	7	0.367
30	14949	6	0.467
30	71827	5	0.567
30	1649373	4	0.700
33	13788017	4	0.727
34	43691	4	0.471
35	2201	7	0.343
35	122921	6	0.486
35	279527	5	0.543
35	15610967	4	0.686
36	2071	9	0.333
36	24309	8	0.417
36	73815	7	0.472
36	959595	6	0.556
36	4740255	5	0.639
36	123818877	4	0.750

Notice that minimum-distance-3 codes are not listed in the above table. The necessary and sufficient condition that minimum distance be over 3, is found to be: 1) exponent of 2, modulo A, is same as code length; 2) the values 1 and -1 are not in the same orbit, i.e., dichotomic. Thus, given any code length e, we look at the smallest divisor A of 2^e-1 that satisfies the above conditions.

We have found also that decoding procedure analogous to Kasami's [12] method is good for the arithmetic codes. But the general decoding scheme when the efficiency is high is still under study.

R. T. Chien
S. J. Hong
F. P. Preparata

11.3. Digital Systems

11.3.1. A Study of a Feedback Time-Sharing System

Many queue disciplines have been studied for use in time-sharing systems. In a first-come-first-serve system, each job joins the end of a queue to wait its turn for service. When it is served, it receives service until it is completed. In the round-robin system, a job joins the end of the queue and awaits service. Service is given to each job one quantum at a time. If the job being served does not finish during that quantum, it returns to the end of the queue to await further service.

Multiple-queue disciplines have been studied in which a job joins a queue of lower priority after each quantum in which it receives service. Before jobs in a queue are given service, all queues of higher priority must be empty.

The first-come-first-served system treats jobs with long service times favorably, but penalizes short jobs, because they must wait until all jobs ahead of them in the system have completed service. The round-robin system is more favorable than the first-come-first-served system for short jobs, but less favorable for long jobs.

A system similar to the first-come-first-served system would be favorable for very short jobs if arriving jobs were given one initial quantum of service and then placed at the end of the queue to await service on a first-come-first-served basis, if they did not complete service in the first quantum. This first-come-first-served system with preemption is more favorable for very short jobs and for very long jobs than the round-robin system is, but is less favorable for jobs of medium length.

If the probability of the arrival of a job during a quantum of length Q, is a constant, p (the result of a Bernoulli trial), and if the number of quanta of service required by a job is chosen independently from the geometric distribution,

$$s_n = (1-\sigma)\sigma^{n-1} \qquad n=1,2,3,\ldots,$$

the expected waiting time, in the first-come-first-served system with preemption, of a job which requires n quanta of service, is given by

$$T_n = Q\left[1 + \frac{n-1}{1-p} + \frac{p\sigma}{(1-p-\sigma)(1-p)(1-\sigma)}\right]$$

except that the value $T_1=Q$ obtains for n=1. In the above expression, the waiting time before the first quantum of service has been neglected.

The first-come-first-served system with preemption may be studied when the swap time, the time required to transfer a job into and out of the server, is not assumed to be zero, but is assumed to be a constant value. The expressions obtained are in closed form but are rather long.

E. J. Beach

11.3.2. Associative Processor

The sheer bulk of inventories, references, and classifications of items in our modern society has made searching for a specific item a difficult task. In many applications, we have been forced to use computers characterized by an addressed memory and a central processing unit; but they are slow when the amount of data to be searched is large. Some effort was made towards designing a processor with a central processing unit and an associative, rather than an addressed, memory. Finally, an associative processor, which is characterized by distributed process control in an associative memory, was proposed by Sturman.

His processor is a linear array of identical cells. In it, a search must be conducted on the basis of a sequence of properties, like the sequence of characters in an English word, rather than on the basis of an unordered set, like the set of properties of a document. The Sturman Processor is a general purpose computer. However, its architecture (i.e., its overall design and method of operation) is not efficient for information retrieval and its relatively complicated queries. The Sturman Processor appears rather inefficient with respect to the distribution of logic between the data storage function and the processing functions.

One of our basic objectives is the identification of an allocation of cost between these two components which could be considered optimal, or quasi-optimal, in a cost-effectiveness analysis.

Preliminary investigation has shown that a modification of one of the ASP processors, designed by the team, D. A. Savitt, H. H. Love, and R. E. Troop, is the most suitable so far considered. While further investigations are being carried out, this evolved model is being studied in some detail.

The basis of this model is a linear array of cells, each containing a fixed-length, relatively-large word, which are connected to a collector channel and a distributor channel. Within the cells, a well-ordered priority structure exists, selecting one cell each time unit for reading out a word, which may contain an instruction or data to be output from the processor, through the collector channel, and then distributing the word to all cells via the distribution channel. A variety of instructions exist to write in a preselected cell, read out of a preselected cell, to preselect a cell, or to perform search operations.

Search operations are obviously most important. Within the linear array of cells, an arbitrary length of contiguous cells stores a phrase. Each phrase may be data, for example, a collection of unordered properties of a document, or a sequence of instructions. Each phrase containing data is stored in cells contiguous to one regular cell acting as a control cell. The bits of the word in the control cell can be used to record matches of the operand field of a search instruction with any word in the phrase

associated with that control cell, or logical functions of several such matches. The control cell is also capable of counting in binary the number of matches and subsequently detecting and recording threshold searches.

An important feature of this processor is its ability to send out instructions to all cells containing data, and simultaneously detect a match in all cells. This parallel operation speeds up the search operation. Further, instructions can be sent out in parallel to each control cell of a data phrase and the results of complex queries can be simultaneously evaluated and recorded. At the end of such a search, a comparatively small number of words are outputted from the processor, thus minimizing the use of the channels, which are the only serial elements of the processor, and are thus the bottleneck to very fast searching.

Attention is now being given to many engineering aspects that have a bearing on the architecture of the processor because previous papers have tended to incorporate highly idealized models of cells. Associated problems in input-output are also being considered.

G. J. Lipovski

11.4. Retrieval Systems

11.4.1. Document Retrieval Based on Semantic and Bibliographic Indexing

Document retrieval based either on coordinate indexing or on citation couplings has been extensively reported in the literature. It is rather clear that each of these devices autonomously suffers from limitations which result into an undesirable reduction of precision and recall in the retrieval operation. To the best of our knowledge, coordinate indexing and

citation couplings have not been exploited in their combined operation. It is felt that the consensus of these two essentially independent structures, semantics and citation, will act as to diminish their respective noises.

In fact, it is well known that lack of consistency in indexing and in query formulation, together with the shallow structure, if any, of coordinate indexing, are detrimental with respect both to relevance and recall. Furthermore, information which may be relevant to a query may appear as a noncentral part of a document and most probably may not be represented in the coordinate indexing description of the document: this would clearly result into loss of recall. It is a reasonable assumption, on the other hand, that documents in a given subject area--and, therefore, candidates for a retrieved set--will exhibit a rather rich interconnection through citation. The exploitation of this structure should help increase the relevance measure of the retrieved set. Moreover, as noted before, inadequately indexed sections of documents which are, however, very relevant to a particular topic are likely to be accessible through citation, thereby increasing the recall measure.

The previous paragraph tacitly implies a fact that has emerged from experience in the implementation of document-retrieval systems, i.e., the superiority of the semantic structure over the citation structure when each is autonomously used (typically the semantic structure is susceptible of revision and betterment, while the citation structure is mechanically extracted from the collection). It is just obvious that the dangers of errors of citation, either through omission, or carelessness (imperfect information of

the author about a reference) or scarce relevance (a reference was very marginally relevant to the topic) are far greater than those deriving from semantic noise. The above discussion, however, asserts the possibility that the analysis of the citation structure will prove to be a powerful auxiliary tool in increasing the performance of the retrieval system.

A pilot experiment, aimed at substantiating the described rationale, is currently under way. In order to make good use of the investigators' background and familiarity with the literature, a small data base is tentatively chosen in the field of computers and computer-aided instruction.

If the results of these preliminary investigations will be encouraging, we plan to establish a sizable data base and to investigate the theoretical aspects of the problem. As a corollary we plan to observe the system deterioration deriving from the substitution of indexer-assigned descriptors with text-generated descriptors. The objective of this investigation is the comparative evaluation of several systems, i.e., systems based on the following retrieval devices:

1) indexer-assigned descriptors, citation
2) text-extracted descriptors, citation
3) indexer-assigned descriptors.

Comparison between 1) and 2) should provide information on the efficacy of carefully controlled indexing in a combined system. Comparison between 2) and 3) should evaluate the power of a system whose retrieval structure is derived in an entirely mechanical fashion.

F. P. Preparata
E. R. Lyman

11.5. Stochastic Systems

11.5.1. Estimators for Optimally Sensitive Systems

The synthesis of the optimally sensitive controller presented in Ref. 13 is modified to include a first-order treatment of measurement noise. The system model is given by

$$\dot{y} = g(y,n,q,t) \quad , \quad y(t_o) = y^o \tag{1}$$

where the notations of Ref. 13 are used. The synthesis procedure is based on the linearization of the optimally-adaptive control $\tilde{u}$ about nominal values y^{o*} and q^*,

$$\tilde{u} = u^* + \omega \Delta y^o + \gamma \Delta q \tag{2}$$

where $\Delta y^o = y^o - y^{o*}$, $\Delta q = q - q^*$. Similarly, a linear approximation is used for the trajectory

$$\tilde{y} = y^* + \zeta \Delta y^o + \sigma \Delta q \tag{3}$$

where y^* is the nominal trajectory. The synthesis procedure is based on estimating Δy^o and Δq from the noisy measurement of $\tilde{y}$, and using these estimates $\widehat{\Delta y}^o$ and $\widehat{\Delta q}$ in (2) to obtain the optimally-sensitive controller.

If the estimation problem is separated from the optimization problem, explicit expressions for the estimators may be derived. The limiting case of zero noise is investigated in particular, and the relation of the results to the degenerate deterministic case is considered.

A. H. Haddad
P. V. Kokotović

11.5.2. Application of a Class of Step Processes

Further progress was made in the application of the step processes considered in Ref. 14 to filtering and interpolation. The performance of some sampling and reconstruction schemes with such processes is investigated. The results will be reported at a later date upon completion.

A. H. Haddad

11.5.3. A Study of a Predictive Quantizing System

This work presents some theoretical properties of a predictive quantizing system for the transmission of random signals. Predictive systems have often been compared to pulse code modulation with the signal-to-quantizing-noise ratio being used as a performance criterion. The errors in this type of analysis are indicated, and the systems are compared on the basis of the mean-squared error criterion. Expressions for the mean-squared error are derived for some processes, and several bounds are also derived.

The Gaussian case is discussed in detail, and it is shown that the predictive system gives no improvement over the conventional communication system. Furthermore, it is shown that the optimum post-filter for Gaussian processes is the inverse of the prefilter. The class of linear Markov processes is also considered. In this case, an example is presented which clearly illustrates the limited region of improvement for the predictive system. The relation of the improvement to the correlation of the input samples is also illustrated.

J. V. Bodycomb
A. H. Haddad

11.6. Information Theory--Coding Theorems for Continuous Channels

The investigation of coding theorems in the sense of Shannon and of explicit codes for time-continuous, amplitude-continuous channels is being pursued. A formulation of the coding theorem for channels in which colored Gaussian noise is added to the output of a filter with square-integrable kernel has been found, but it is applicable only to a severely restricted class of filters and Gaussian noise. It is hoped to extend these results by approximating channels not satisfying these restrictions with channels that do. Proofs of coding theorems for these channels appear to depend upon relating properties of the transfer function of the filter and the spectral density of the noise to properties of the channel.

R. B. Ash
J. K. Moore

11.7. Communication Technology

11.7.1. Resolvers Using Modulated Integrators

An electronic resolver is an analog computing element whose input may be regarded as the analog of a shaft rotation φ and whose outputs are the analogs of the sine and cosine of the rotation, $\sin\varphi$ and $\cos\varphi$. Usually, it is the rotation rate $\dot{\varphi}$ that is taken to be the input.

Realizations, such as shown in Fig. 11.1, using electronic integrators are well known. Since the rate of change of $\sin\varphi$ is

$$\frac{d}{dt}\sin\varphi = \dot{\varphi}\cos\varphi, \tag{1}$$

while that of $\cos\varphi$ is

$$\frac{d}{dt}\cos\varphi = -\dot{\varphi}\sin\varphi, \tag{2}$$

then it is the right-hand side of (1) that is to be formed at the output of a multiplier and supplied to an integrator to form $\sin\varphi$, except that for polarity-inverting integrators, the usual sort, it is the negative of $\dot{\varphi}\cos\varphi$ that is to be supplied, as shown for the upper integrator of Fig. 11.1. Similarly, the lower integrator is to have the input which is the negative of the right-hand side of Eq.(2), $\dot{\varphi}\sin\varphi$.

Apart from obvious uses in analog-computer simulation of mechanical systems, there are potential applications of such resolvers as frequency-controllable oscillators and in frequency modulators for fixed-frequency oscillators. As a frequency-controllable oscillator, it is the input $\dot{\varphi}$ that represents the controlled frequency, for if this be set to be the constant value ω_o, then the output oscillations have the waveforms represented by $\sin\omega_o t$ and $\cos\omega_o t$. The use in a frequency modulator is indicated in Fig. 11.2. There the multiplication of $\sin\varphi$ by $\sin\omega_o t$ and the subtraction of that from the product of $\cos\varphi$ with $\cos\omega_o t$ fulfills the well-known trigonometric identity

$$\cos(\omega_o t+\varphi) = \cos\omega_o t\cos\varphi - \sin\omega_o t\sin\varphi, \tag{3}$$

and frequency modulation obtains because the instantaneous frequency of the oscillation represented on the left of Eq.(3) is the rate of change of $\omega_o t+\varphi$, namely $\omega_o+\dot{\varphi}$, in which $\dot{\varphi}$ is the instantaneous frequency deviation. A

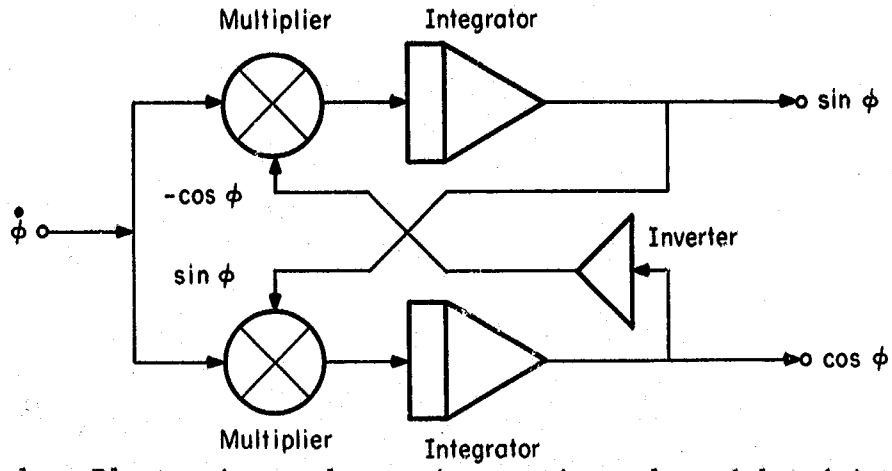


Fig. 11.1. Electronic resolver using continuously-modulated integrators.

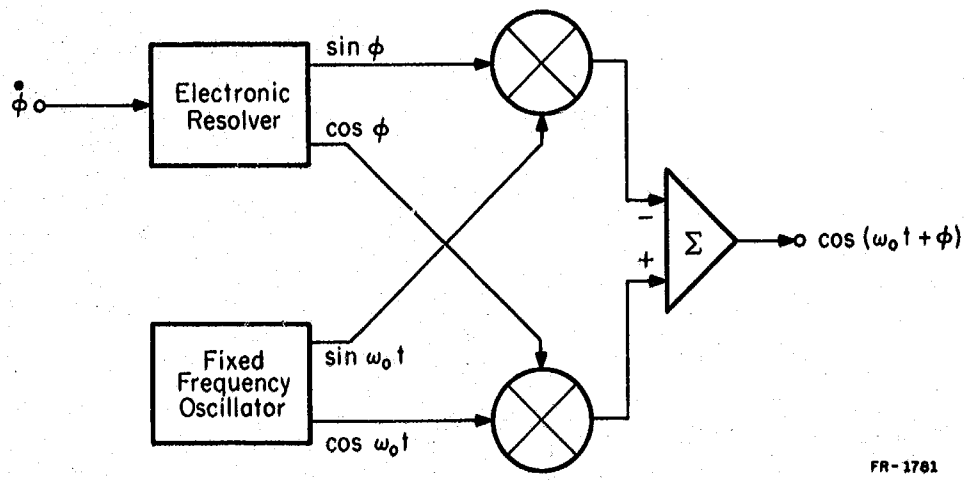


Fig. 11.2. Use of electronic resolver in a frequency modulator.

(OVERLEAF BLANK)

simplified version of the system of Fig. 11.2 is the well-known Armstrong modulator based on the approximation

$$\cos(\omega_o t+\varphi) \approx \cos\omega_o t-\varphi\sin\omega_o t, \tag{4}$$

valid for $\varphi<<1$.

Because of the difficulties in constructing and maintaining sufficiently stable integrators and sufficiently accurate multipliers, the potential applications of resolvers of the kind shown in Fig. 11.1 have not been widely exploited. An alternative resolver-like system, described below, is being studied in the hope that a new approach will offer means of alleviating such difficulties.

11.7.2. Resolvers Using Switched-Polarity Integrators

An arrangement similar to that shown in Fig. 11.1, in which the analog multipliers are replaced by polarity-switching gates (PSG), can provide a resolver-like operation, except that the sinusoidal dependence on φ is also replaced by a serrasoidal dependence. There are two kinds of serrasoids involved, namely $\mathrm{saw}\varphi$ and $\mathrm{caw}\varphi$ as shown in Fig. 11.3. The rate of change of $\mathrm{saw}\varphi$ is readily seen to be

$$\frac{d}{dt}\,\mathrm{saw}\varphi = \dot{\varphi}\,\mathrm{sgn}(\mathrm{caw}\varphi), \tag{5}$$

while that for $\mathrm{caw}\varphi$ is

$$\frac{d}{dt}\,\mathrm{caw}\varphi = -\dot{\varphi}\,\mathrm{sgn}(\mathrm{saw}\varphi), \tag{6}$$

in which the signum function, sgn( ), is defined to have the value +1 for positive values of its argument and to have the value -1 otherwise. The PSG's shown in Fig. 11.4 transmitting $+\dot{\varphi}$ for positive values of the control signal, and $-\dot{\varphi}$ otherwise, provide the multiplication by the signum function so that the arrangement of Fig. 11.4 obeys Eqs.(5) and (6). True resolver operation requires the conversion of the serrasoidal waveforms to sinusoidal waveforms, of course, but this conversion may be provided by a pair of memoryless nonlinear devices each exhibiting an input-output transfer curve which is a sinusoid for inputs in the range from $-\pi/2$ to $\pi/2$.

There are three main areas of investigation that could bear on demonstrating the utility of serrasoidal resolvers. Two of these have to do with hardware questions, whether (a) the replacement of multipliers by PSG's can introduce sufficient simplicity to justify the necessity of using serrasoidal-to-sinusoidal converters, and whether (b) means of controlling integrator drift can be simpler or more efficacious in the serrasoidal version. The third has to do with systems questions, namely whether (c) the serrasoidal-to-sinusoidal converter may be eliminated in FM-like communication systems while retaining many of the advantages of FM. Favorable results from such studies would place the serrasoidal resolver in an important position in applications in which wide-range, stable, frequency or phase modulation would be used.

A preliminary study, expected to lead to a master's thesis, has been undertaken in which a serrasoidal resolver, using output-waveform zero-crossings to activate a resetting function for control of integrator

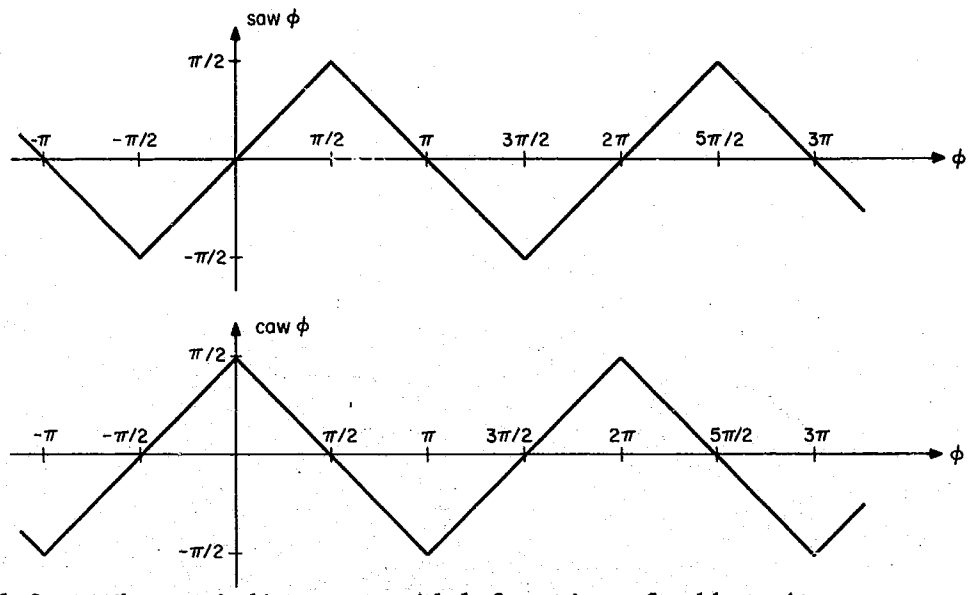


Fig. 11.3. The periodic serrasoidal function of odd parity, saw φ, and its complement of even parity, caw φ.

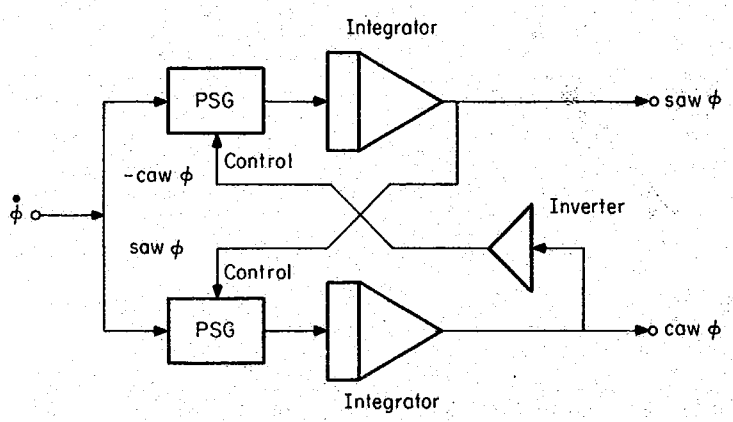


Fig. 11.4. Electronic resolver using switched-polarity integrators.

(OVERLEAF BLANK)

drift, is to be designed and constructed. This work is nearing completion, and it is expected to provide a basis for deciding the extent to which the more ambitious investigations cited above should be pursued.

D. H. Cooper
R. J. Kubala

11.8. References

1. A. George Schilinger, "A class of Solid-Burst-Error-Correcting Codes," Research Report PIBMRI-1223-64, Polytechnic Institute of Brooklyn, Brooklyn, New York (April 1964).

2. A. D. Wyner, "A Note on a Class of Binary Cyclic Codes Which Correct Solid-Burst Errors," IBM Journal of Research and Development (January 1964).

3. C. M. Melas, "A Cyclic Code for Double Error Correction," IBM Journal of Research and Development 4, 364 (1960).

4. A. B. Fontaine and R. G. Gallager, "Error statistics and coding for binary transmission over telephone circuits," Proc. IRE 49 (June 1961).

5. R. T. Chien, D. T. Tang, E. S. Barrekette and A. M. Katchev, "Analysis and improvement of Photostore error rates," Proc. IEEE 56 (May 1968).

6. I. S. Reed and A. Solomon, "Polynomial codes over certain finite fields," Journal SIAM 8 (June 1960).

7. A. W. Nordstrom and J. P. Robinson, "An Optimum Nonlinear Code," Inf. and Control 11, 613 (1967).

8. F. P. Preparata, "Weight and Distance Structure of Nordstrom-Robinson Quadratic Code," Inf. and Control (to appear) (1968b).

9. S. J. Hong, "On Minimum Distance of Multiple Error Correcting Codes," Jan. 1967, CSL Report R-336.

10. R. T. Chien, S. J. Hong, F. P. Preparata, "Some contributions to the theory of Arithmetic Codes," Proc. Hawaii Int. Conf. on System Sciences, Jan. 1968.

11. R. T. Chien, S. J. Hong, F. P. Preparata, paper in preparation.

12. T. Kasami, "A Decoding Procedure for Multiple-Error-Correcting Cyclic Codes," IEEE Trans. on Inf. Theory IT-10, 134 (April 1964).

13. P. V. Kokotović, J. B. Cruz, J. E. Heller, and P. Sannuti, "Synthesis of Optimally Sensitive Systems," Proc. IEEE, Vol. 56 (August 1968).

14. A. H. Haddad, "On a Class of Step Processes," Proc. 2nd Annual Princeton Conference on Information Sciences and Systems, pp. 169-173 (March 1968).

12. NETWORKS[†]

T. J. Aprille, Jr.
W. Mayeda*
S. Toida
T. N. Trick
N. Wax

12.1. An Application of Computer Optimization Techniques to Network Synthesis

Computer optimization techniques were applied to two specific network synthesis problems. The first problem was, given an initial circuit, find values for the elements of the circuit so that some characteristic of the resulting circuit could "best match" in some sense a given desired characteristic. It was found that this method of design was not very successful. A good initial guess was necessary in order to avoid local minima which resulted in solutions which were obviously poor designs.

The second problem involved using the optimization technique as an intermediary step in a practical synthesis problem. The procedure is outlined as follows: Realize a given network characteristic by an ideal element circuit. This could be done by any means one wishes; for example, in a filter problem it could be accomplished by using the lossless realization. Now replace the ideal elements by practical elements and use the optimization technique to bring the resulting practical network characteristic back to the desired characteristic. Several examples of each approach are given in a CSL report which will be issued soon.

[†]This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR 931-67 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

*On leave of absence as Visiting Professor at the University of Texas at Austin.

One of the examples was to practically realize a lowpass filter with cutoff frequency of 1 and magnitude equal to 0.5. A ninth-order "ideal" Chebyshev realization of this is shown in Fig. 12.1. Then, to simulate a practical circuit built from this schematic, resistances were added in series with each coil. The resistance value added was equal to 0.1 times the value of the coil. The resulting initial transfer magnitude curve is shown as "lossy characteristic" in Fig. 12.2. An optimization technique was applied to the circuit using a mean-square-error criterion and the weighting indicated in Fig. 12.2. The new parameter values are shown in parentheses in Fig. 12.1; the resulting transfer magnitude curve is given as "optimal characteristic" in Fig. 12.2.

This final curve has a high ripple in the pass band. To improve this, a different weighting was used in the optimization procedure. This again changed the values of the resulting circuit and led to the transfer magnitude curve labeled "different weight" in Fig. 12.2. Weighting was found to be very important in this example. A bandpass example is also given in the report to further stress the importance of weighting in obtaining a good solution. In the report, the many different optimization techniques that could be used in these examples are also summarized.

T. J. Aprille, Jr.
T. N. Trick

12.2. Stability of Nonlinear Networks

Asymptotic stability has been proven for a class of networks containing a nonlinear time-varying capacitor whose incremental capacitance is greater than zero. The network considered is shown in Fig. 12.3.

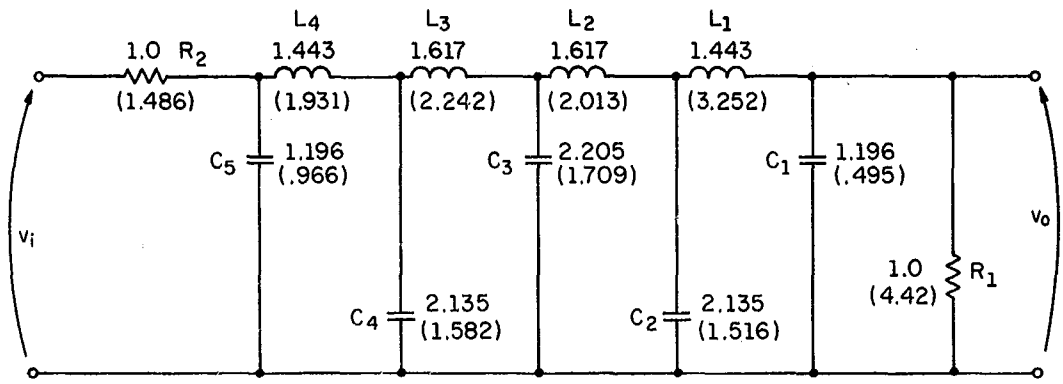


Fig. 12.1. Ninth-order Chebyshev low-pass filter.

(OVERLEAF BLANK)

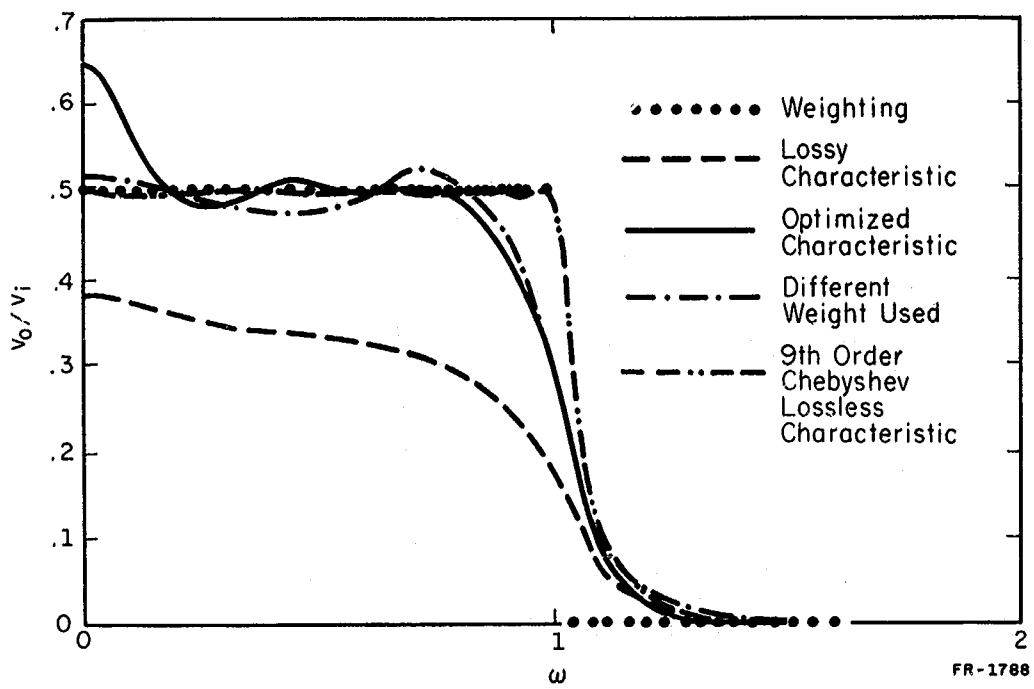


Fig. 12.2. Magnitude of the transfer function.

(OVERLEAF BLANK)

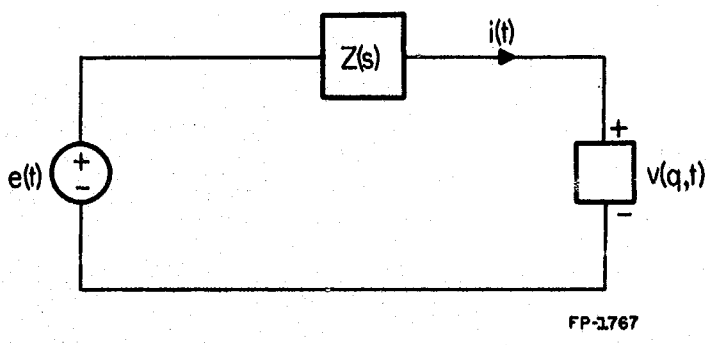


Fig. 12.3. Network containing a nonlinear capacitor.

(OVERLEAF BLANK)

It is assumed that $e(t)=e_p(t)-e_t(t)$ is bounded, and that $e_p(t)$ is the steady state Thevenin voltage and that $e_t(t)$ is the transient component which asymptotically approaches zero as $t \to \infty$. Also, we assume that

$$0 < \alpha \leq \frac{v(q_1,t) - v(q_2,t)}{q_1 - q_2} \leq \beta, \text{ for all } q \text{ and } t,$$

so that

$$H(s) = \frac{1}{sZ(s) + \frac{1}{2}(\alpha+\beta)}$$

is the Laplace transform of an absolutely integrable function. These conditions placed on the network can be satisfied in most instances.

Theorem: The above network is asymptotically stable if the step response $i_1(t)$ indicated in Fig. 12.4 is greater than or equal to zero for all $t \geq 0$.

This result implies that if $e_p(t) = e_p(t+T)$, then the current $i(t)$ in Fig. 12.3 asymptotically approaches a unique, bounded, steady-state solution obeying $i_p(t) = i_p(t+T)$. Thus, no subharmonic oscillations or other anomalous nonlinear phenomena will occur. It is interesting to note that if $Z(s)$ is the impedance of an RC network, then the conditions of the theorem are always satisfied. These results have been submitted for publication.

T. N. Trick

12.3. The Characterization of Biplanar Graphs

Several new theorems have been established on the characterization of biplanar graphs. These theorems are given below.

Definition 1: A regular graph of degree n is a graph all of whose

vertices have n edges incident to them. For example, if n=4, then every vertex of the graph has four edges incident to it.

Theorem 1: A regular graph of degree 4 is biplanar.

Definition 2: A regular graph with 14 vertices of degree 7 will be referred to as $K_{7,7}$. In this graph every pair of an even vertex and an odd vertex is connected by an edge.

Theorem 2: The graph $K_{7,7}$ is not biplanar.

Theorem 3: If we remove one edge (any one) from $K_{7,7}$, then the remaining graph is biplanar.

S. Toida

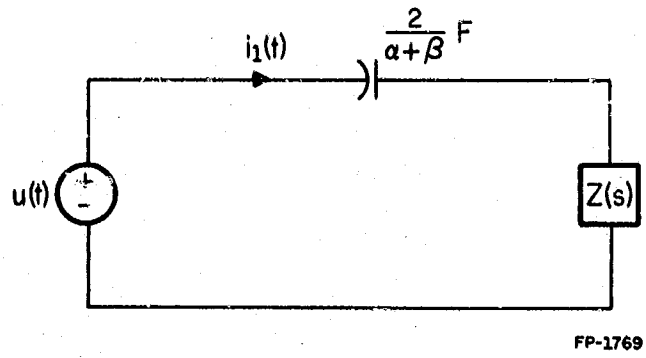


Fig. 12.4. Determination of asymptotic stability.

(OVERLEAF BLANK)

13. CONTROL SYSTEMS[†]

J. B. Cruz, Jr.
C. I. Chen
J. Dalmia
G. D'Ans
J. E. Heller
R. Hecht
M. Jamshidi
P. Kokotović
J. Medanić
W. R. Perkins
P. Sannuti
G. Schumacher
R. Stefanek
D. F. Wilkie

13.1. Introduction

This group is concerned with the analytical, computational, and simulation aspects of control system analysis and design. During this reporting period, major emphasis has been given to a detailed investigation of control in the presence of uncertainties. Generation of sensitivity functions, singular perturbation techniques, design of approximately optimal systems, use of game theory, and minimax design have received attention. Simulation studies of two systems, a manned orbital space station and a cement kiln, have been carried out.

J. B. Cruz, Jr.
W. R. Perkins

13.2. Use of the Companion Transformation in Parameter Optimization and Adaptive Control

The transformation to companion canonic form for single-input, linear, time-invariant systems has been shown to have important applications in the sensitivity analysis of control systems as well as in parameter optimization and model reference adaptive control problems. Two previously unknown properties of a system in companion form have been demonstrated.

[†]This work was supported by the Air Force Office of Scientific Research under Grant No. AFOSR 931-67 and AFOSR 68-1579 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

These are the Total Symmetry Property of the sensitivity matrix of a system in companion form and the Complete Simultaneity Property. The latter states that all the sensitivity functions of the states of a companion-form system can be generated by linear combinations of the signals on one sensitivity model of the system and the system states. Using these properties, a method is developed to generate by one n^{th} order sensitivity model all the sensitivity functions $(\partial x_i/\partial v_j)_{\underset{\sim}{v}o}$, $i = 1,\ldots,n$, $j = 1,\ldots,r$ of the state of any single input, linear, time-invariant, controllable n^{th} order system which depends on r different parameters. This represents an improvement over known methods for generating the sensitivity functions, which generally require a composite dynamic system of order n(r+1). It was demonstrated that use of this result can yield time savings in the sensitivity analysis of systems and in adaptive control problems. This result was then extended to multi-input normal systems. A study was also made of the effect of performing parameter optimizations in the space of the coefficients of the characteristic equation as opposed to the original parameter space.

Finally, a new approach for the design of model reference adaptive control for single-input, linear, time-invariant plants was developed. The performance index minimized was the norm of the difference between the model and system companion transformations, with the constraint that the system and model eigenvalues are the same. This constrained minimization was converted to an unconstrained algebraic minimization by including state feedback in the controller. Since no iterative solutions of the differential equations are required, the solution time is less than that required for

previous approaches to the model reference adaptive problem involving minimization of an integral performance index.

This project was completed in July, 1968, and a report R-386, "Use of the Companion Transformation in Parameter Optimization and Adaptive Control," by D. F. Wilkie was issued in July. One paper, "Essential Parameters in Sensitivity Analysis," based on part of the report was presented at the Second IFAC Congress on Systems Sensitivity and Adaptivity in Dubrovnik, Yugoslavia in August, 1968. Other papers based on the report have been submitted for publication.

D. F. Wilkie
W. R. Perkins

13.3. Singular Perturbation Method in the Theory of Optimal Control

A detailed description of physical phenomena often requires a mathematical model of high order. The existing optimal control synthesis procedures generally become impractical for dynamic systems of high order. One is therefore forced to construct a simplified low-order model in order to make existing numerical procedures applicable. In constructing a low-order model it is a common practice to neglect some small time constants, masses, inductances, capacitances and similar "parasitic" parameters. Typically, a control is synthesized for such a low-order model and then its applicability is checked on a more realistic high-order model. Unfortunately, more frequently than not the result of such a low-order design is unsatisfactory.

By the proposed "optimally sensitive design," the optimal control

for a high-order model of the plant is approximated by some functions obtained from two low-order models, the second being the sensitivity model of the first. In the first step of the method, an (n+m)-th order model of the plant is constructed in such a way that it reduces to an n-th order model when a plant parameter λ attains a "critical" value λ_c, (usually $\lambda_c = 0$). The n-th order model with $\lambda = \lambda_c$ is taken as an approximation of the (n+m)-th order model in which the value of λ is perturbed from the critical one, $\lambda \neq \lambda_c$. Then the idea of a neighboring optimum or optimally sensitive control is employed, and a control approximately optimal in some neighborhood $\delta\lambda = \lambda - \lambda_c$ of λ_c is obtained. Since such a "singular perturbation" $\delta\lambda$ causes the model order to become n+m, the same control also approximates the control optimal for the more accurate high-order model. In this way, an optimally sensitive control is achieved for the case when the optimal sensitivity is required with respect to the change of model order.

There are mainly two problems to be solved in developing the "optimally sensitive design." Firstly, it must be shown under which conditions such a method is possible. Secondly, the algorithm of such a design procedure should be shown to be computationally simpler than the design of the high-order model. For a class of nonlinear systems, the first problem is solved by using the singular perturbation theory of ordinary differential equations. Several examples illustrate the practical applicability and the computational simplicity of the method.

This project was completed in June, 1968, and is reported in Coordinated Science Laboratory Report R-379.

P. Sannuti
P. Kokotović

13.4. An Approximation Theorem for Linear Optimal Regulators

The technique of Taylor series expansion is a convenient tool for sensitivity analysis and near optimum design of dynamical systems depending on parameters [1,2]. The theorem presented below establishes a property of expansions of optimal feedback matrices for linear state regulator systems with quadratic performance indices. It extends a recent theorem for expansions of optimal controls [2]. Since the theory of linear systems with quadratic performance indices is widely used in optimal control, estimation, and Lyapunov stability, the theorem has a variety of applications.

Consider the well-known linear state-regulator problem [3] for the plant

$$\dot{x} = Ax+Bu \tag{1}$$

with performance index

$$J = \tfrac{1}{2}\int_{t_o}^{T} (x'Qx+u'Ru)dt, \tag{2}$$

where x is the n-dimensional state vector and u is the r-dimensional control vector, dot denotes differentiation with respect to time, and a prime denotes a transpose. Let the matrices A, B, Q, and R be analytic functions of a scalar parameter $\alpha = 0$. These matrices may be time-varying,

but they must satisfy the usual assumptions of the linear state regulator problem [3].

The optimum matrix K used in the state feedback regulator

$$u = R^{-1}B'Kx \tag{3}$$

is an analytic function of α, $K = K(t,\alpha)$. This follows from the theorem on differentiability of solutions of ordinary differential equations with respect to a parameter when applied to the Riccati equation

$$\dot{K} = -KA-A'K+KSK-Q,\ K(T,\alpha) = 0, \tag{4}$$

where $S = BR^{-1}B'$.

Let $M = M(t,\alpha)$ be a matrix defined by

$$M(t,\alpha) = \sum_{i=0}^{m-1} \frac{\alpha^i}{i!} \left.\frac{\partial^i K(t,\alpha)}{\partial \alpha^i}\right|_{\alpha=0} \tag{5}$$

The problem considered here is to compare the performance J_M of the system

$$\dot{x} = (A-SM)x \tag{6}$$

with the state feedback regulator

$$u = -R^{-1}B'Mx, \tag{7}$$

to the performance J_K of the system

$$\dot{x} = (A-SK)x, \tag{8}$$

with the optimum state feedback regulator (3).

Theorem: The first 2m terms of the Taylor series for J_M match the first 2m terms of the Taylor series for J_K,

$$\frac{\partial^i J_M}{\partial \alpha^i}\bigg|_{\alpha=0} = \frac{\partial^i J_K}{\partial \alpha^i}\bigg|_{\alpha=0}, i = 0,1,\ldots,2m-1 \tag{9}$$

Proof: The expression for J_K is well known to be [3]

$$J_K = \tfrac{1}{2}x'Kx\big|_{t-t_o}. \tag{10}$$

Similarly it can be shown (see for example, Ref. 4, pp. 81-84) that the performance J_M is

$$J_M = \tfrac{1}{2}x'Px\big|_{t-t_o}, \tag{11}$$

where $P = P(t,\alpha)$ is the solution of the linear symmetric matrix equation

$$\dot{P} = -P(A-SM)-(A'-MS)P-MSM-Q, \qquad P(T,\alpha) = 0. \tag{12}$$

Define Γ as

$$\Gamma = P-K. \tag{13}$$

Then to prove (9) it is equivalent to prove that at $\alpha = 0$ and $t = t_o$, that Γ and its first 2m-1 derivatives are zero matrices,

$$\frac{\partial^i \Gamma(t,\alpha)}{\partial \alpha^i}\bigg|_{\substack{t=t\\ \alpha=0}} = 0, \; i = 0,1,2,\ldots,2m-1. \tag{14}$$

From (5) there obtains

$$K-M = \alpha^m \Lambda, \tag{15}$$

in which

$$\Lambda = \sum_{i=m}^{\infty} \frac{\alpha^{i-m}}{i!} \frac{\partial^i K(t,\alpha)}{\partial \alpha^i}\bigg|_{\alpha=0}. \tag{16}$$

Subtracting (4) from (12) and using (13) and (15), there obtains

$$\dot{\Gamma} = -\Gamma G - G'\Gamma - \alpha^m(\Gamma S\Lambda + \Lambda S\Gamma) - \alpha^{2m}\Lambda S\Lambda, \ \Gamma(T,\alpha) = 0, \tag{17}$$

where $G = A-SK$. This equation shows that (14) holds not only at $t = t_o$ but also for all $t\in[t_o,T]$. To see this, first check (14) for $i = 0$. That is, find the solution $\Gamma(t,\alpha)$ of (17) for $\alpha = 0$. At $\alpha = 0$, (17) reduces to a homogeneous linear equation

$$\dot{\Gamma} = -\Gamma G - G'\Gamma, \ \Gamma(T,0) = 0, \tag{18}$$

whose solution $\Gamma(t,0)$ must be zero for all t, since it is zero at $t = T$. Next, denote derivatives by subscripts, $\Gamma_i = \partial^i\Gamma/\partial\alpha^i$, $G_j = \partial^j G/\partial\alpha^j$, etc., $i,j = 0,1,2,\ldots$. To verify (14) for $i = 1$, differentiate (17) with respect to $\alpha = 0$,

$$\dot{\Gamma}_1 = -\Gamma_1 G_o - G'_o\Gamma_1 - \Gamma_o G_1 - G'_1\Gamma_o - m0^{m-1}(\Gamma_o S_o\Lambda_o + \Lambda_o S_o\Gamma_o), \ \Gamma_1(T,0)=0, \tag{19}$$

where $m0^{m-1}$ is 1 when $m = 1$ and is 0 when $m > 1$. Since the result of the verification for $i = 0$ is $\Gamma_o = 0$, (19) becomes homogeneous and hence

$\Gamma_1 = 0$. Using induction, verify whether the assumption of $\Gamma_j = 0$, for $j = 0,1,\ldots,i$, makes the equation for Γ_{i+1} homogeneous, and determine the maximum i for which this is true. Differentiating (17) i+1 times and letting $\alpha = 0$, there obtains

$$\dot{\Gamma}_{i+1} = -\Gamma_{i+1}G_o - G'_o\Gamma_{i+1} - F, \quad \Gamma_{i+1}(T,0) = 0. \tag{20}$$

For $i < 2m-1$, each term in F is linear in some Γ_j, $j = 0,1,\ldots,i$. Since $\Gamma_j = 0$, for $j = 0,1,\ldots,i$, there obtains $F = 0$, and the linear equation (2) is homogeneous. Thus $\Gamma_{i+1}(t,0)$ is zero because it is zero at $t = T$. For $i \geq 2m-1$, it is seen that F may contain a nonzero term because the (i+1)-th derivative of $\alpha^{2m}\Lambda S\Lambda$ may be nonzero. Thus for $i \leq 2m-1$, there obtains $\Gamma_{i+1} = 0$. This proves that (9) holds for $i = 0, 1,\ldots,2m-1$.

The main content of the theorem is in establishing the order to which the system performance approximates the optimal performance, if, instead of the optimum gain matrix, its truncated series M is used. It is sufficient that the order of the truncated series M be one-half of the desired order of the approximation of the optimal performance. Moreover, the parameter α can be selected to make the system dimensionality lower when $\alpha = 0$, than when $\alpha \neq 0$ (singular perturbation of α [5,6]). Although in this case some elements of the system matrices A and B have a pole at $\alpha = 0$, it can be shown [7] that under mild conditions, K and P are analytic in α at $\alpha = 0$. Therefore, the theorem of this paper extends to that class of "singular perturbation" problems.

P. V. Kokotović
J. B. Cruz, Jr.

13.5. Near Optimal Control of Linear Systems with Bounded Control and Minimum Fuel Index

The control for a linear plant with a bounded control and a fixed time fuel index, a given initial state, and a fixed terminal state is well known to be a dead-zone function. For a given initial state the control is piece-wise constant with levels at +1, 0 and -1, and the switching times may be computed by any standard iteration method.

Near optimal control may be obtained for a neighborhood of initial conditions by a perturbation of the switching times from a nominal set of switching times for a nominal initial state. The new switching times are approximated by the nominal switching times plus the partial derivative (sensitivity) of the optimal switch time with respect to initial state, multiplied by the change in initial state. The switch-time sensitivities are computed under the condition that the state sensitivity at the terminal time is zero, i.e., the terminal condition is satisfied to first order in the initial state deviation.

From the numerical examples that were considered, near optimal control required fuel close to the minimum with only a small violation of the terminal constraints for a small neighborhood of the initial state. On the other hand, open-loop control requires a constant fuel for the same neighborhood of initial conditions, but the violation of terminal constraints is worse. For some range of initial states, the open-loop fuel is less than the fuel for the approximately optimal control, thus providing a trade-off for the fuel consumption and satisfaction of terminal constraints. Details

may be found in CSL Report R-384, "Use of Switch Time Sensitivity to Provide Near Optimal Control," by Stefanek and Cruz, June, 1968.

R. G. Stefanek
J. B. Cruz, Jr.

13.6. Performance Segment Design of Linear Control Systems

The Performance Segment Design of Linear Control Systems (see Progress Report for September 1967-February 1968) has been extended to the class of problems with the constraint on disturbances independent of state variables. The problem has been analyzed and the design procedure based on a gradient method has been applied. The results obtained from examples indicate again that this design procedure results in controls giving better performance than that resulting from other controls such as the minimax type control of the nominal optimal control.

J. Medanić
C. Chen

13.7. Singular Solutions and Differential Games

A desire to apply the ideas and techniques of differential games to sensitivity problems and problems of control with uncertainties made necessary a literature survey into this field of control theory. This in turn has led to the formulation of several research projects within the field of differential games. One of these projects concerns the role and significance of singular solutions in differential games. The problem has been considered in detail only by Isaacs (Differential Games, John Wiley, 1965). New results concerning necessary conditions for singular solutions,

the classification of singular controls by rank and degree and interpretation of these results in the game situation leads to a deeper understanding of the problem. A number of pursuit and evasion examples exhibiting a diversity of singular solutions along "universal" surfaces support interest in this problem.

C. Chen
J. Medanić

13.8. Elimination Method in the Solutions of Minimax Problems

The elimination method (see Progress Report for September 1967-February 1968, sec. 13.10) has been developed into an algorithm suitable for computer application. Since the elimination method assures convergence to the minimax with arbitrary selection of nominal points in the reduced domain the efforts in constructing the algorithm were concentrated in reducing computational requirements. The algorithm consists of the following:

i. select a nominal point $c^n \in C_{ni}$

ii. maximize with respect to $p \in P_i$

iii. find the contour gradient and thereby eliminate a portion of C_n, and

iv. select new nominal point in the reduced domain.

The improvements in the method consist in developing a simple way of selecting nominal points (step iv) as well as in showing and applying the fact that maximization (step ii) need not be performed in each elimination. A paper reporting the details has been submitted for publication.

J. Medanić

13.9. Optimal Control of Systems with Unknown of Varying Parameters

The design of a controller for a dynamic system $\dot{x} = f(x,u,t,v)$, where x is the state vector, u is the control vector and v is the unknown or varying parameter vector, is being considered when the performance of a control vector u(t) for a given v(t) is given by a scalar functional S(u,v). A pessimistic approach is taken which consists of determining a particular control $u^*(t)$ which attains

$$\min_{u \in U} \max_{v \in V} S(u,v)$$

where U and V are the respective admissible domains.

The development of a computational algorithm to obtain u^* for non-saddle point solutions is proceeding at this time with promising results.

J. E. Heller

13.10. Synthesis of Suboptimal Nonlinear Regulators

Consider a nonlinear system

$$\dot{x}(t,\epsilon) = Ax(t,\epsilon) + Bu(t,\epsilon) + \epsilon f(x,u) + d(t)$$

$$x(0,\epsilon) = x(t_o),$$

where ϵ is a parameter, x is a state vector, u is a control vector, d is a disturbance vector, A and B are constant matrices, and f is a continuous, differentiable function of x and u. The performance index is

$$J = \tfrac{1}{2}\left[x'(t_f)\, Mx(t_f)\right] + \tfrac{1}{2}\int_{t_o}^{t_f} (x'Qx + u'Ru)\,dt$$

where Q and R are positive definite matrices.

A method has been developed for obtaining the optimal control as a series in ϵ. The determination of the coefficients in the series involves solving linear equations only. It has been shown also that a semi-feedback realization using an nth order approximation involves the same feedback gain for any n including n=1. Only the open-loop portions of the control are changed. The theory has been applied to a fourth order cold-strip mill plant.

M. Jamshidi

13.11. Spacecraft Attitude Control Using Feedback Parameter Optimization

For a Manned Orbital Space Station (MOSS) to have an artificial gravity, it is required to rotate at a constant rate about one axis and have no angular rates about the remaining two axes. Various disturbance torques present cause small changes from these desired rates.

To obtain the desired control, twin gyrotorquer units were chosen as the control devices. The required nonlinear equations were developed to simulate the space station dynamics, the gyrotorquer dynamics and the controller. The structure of the controller was fixed with only the values of various parameters left to be found. Numerical optimization techniques were used to determine the proper feedback parameters. A performance index was defined which would be at a minimum when the proper set of feedback parameters were determined. Several cases of initial conditions and disturbance torques were used which lead to some interesting response characteristics.

G. C. Schumacher

13.12. Simulation of a Wet Process Cement Rotary Kiln

Computer simulation is an important aspect of the study of large scale systems. In this project a mathematical model of a cement kiln was developed, and a computer simulation of this model was made to obtair important temperature and concentration profiles.

The manufacture of cement involves mixing of certain calcareous and silicious materials and sintering them at a high temperature to produce a material with cementing properties. According as the mixture of calcareous and silicious materials (called slurry) contains or does not contain water, the process is wet or dry. The kiln, a combination of a dryer, calciner and reactor, is used to sinter the slurry at high temperature (about $200^{o}C$) and the product obtained is called clinker. This is mixed with gypsum in fixed proportions to produce cement.

The technology of the process in the rotary kiln is based on the principle of heat exchange by counterflow of solids and gas. A cooler at the discharge end of the kiln is used to cool the hot clinker by secondary air which is forced through the clinker bed by a fan. This preheats the secondary air and results in heat economy of the kiln. Various measurements of temperatures and pressures at different points along the kiln length are made to maintain proper temperature gradient and suction of the flue. The kiln is divided into drying, calcining, sintering and cooling zones, the length of which depend upon concentrations of the solid constituents, temperature of the solids and gas phases, etc.

A mathematical model of the kiln can be obtained by selecting

independent and dependent variables, and using material and heat balance relationships to formulate partial differential equations.

Simulations of these equations were undertaken for both steady-state and transient conditions. For steady-state simulation, the dynamic equations, which for the kiln are first order, nonlinear partial differential equations, are reduced to ordinary differential equations with position along the kiln length as the only independent variable. This can be done by letting $\partial/\partial t \to 0$ as $t \to \infty$.

For transient analysis, the kiln is divided into a convenient number of small sections of suitable but equal lengths. Then, for each section and each variable, generalized differential-difference equations are written. The objective is to determine the boundary shifts of the zones in the kiln, defined by the results of the steady-state analysis, when a step input is given to the kiln rotational speed. It should be mentioned that in both steady-state and transient analysis the differential equations were solved on the Control Data 1604 digital computer using the numerical method of Runge, Kutta, and Gill.

Details of this simulation, including complete equations and concentration and temperature profiles, will appear in a forthcoming CSL report.

J. H. Dalmia

13.13. References

1. H. J. Kelley, "An Optimal Guidance Approximation Theory," IEEE Trans. on Automatic Control, AC-9, 375-380 (1964).

2. R. A. Werner and J. B. Cruz, Jr., "Feedback Control which Preserves Optimality for Systems with Unknown Parameters," IEEE Trans. on Automatic Control, AC-13, in press (1968).

3. R. E. Kalman, "Contribution to the Theory of Optimal Control," Bol. Sociedo Mate. Mex., 102-119 (1960).

4. J. LaSalle and S. Lefschetz, Stability of Liapunov's Direct Method, Academic Press, New York, 1961.

5. A. N. Tikhonov, "On the Dependence on Solutions of Differential Equations on a Small Parameter," Mat. Sbornik 22(64), 193-204 (Moscow, 1948). See also, A. N. Tikhonov, "Systems of Differential Equations Containing Small Parameters Multiplying Some of the Derivatives," Mat. Sbornik 31(73), 575-586 (Moscow, 1952).

6. A. B. Vasileva, "Asymptotic Behavior of Solutions to Certain Problems Involving Nonlinear Differential Equations Containing a Small Parameter Multiplying the Highest Derivative," Russian Mathematical Surveys, 18, 13-81 (1963).

7. P. Sannuti and P. Kokotović, "An Approximate Design of Optimal Regulators for High-Order Linear Plants," Preprints IFAC Symp. on System Sensitivity and Adaptivity (Dubrovnik, Yugoslavia, August, 1968).

(OVERLEAF BLANK)

14. SWITCHING SYSTEMS[†]

G. Metze
E. Davidson
T. Gaddess
J. Hayes
R. Isenhart
T. Powell
D. Schertz
S. Seth
A. Wojcik

14.1. NAND Decomposition of Combinational Logic

The development of a NAND decomposition algorithm has been concluded with the extension of the algorithm to permit the specification of network constraints. A fan-in and a fan-out limit for the NAND gates, a separate fan-out limit for variables, and a level limit for each required function can be specified. Further constraints and cost criteria are compatible with the algorithm.

The use of constraints, particularly the level constraint, normally has a "pruning" effect on the solution tree generated by the algorithm and thereby enhances the convergence of the algorithm. For example, the seventy-seven 3-variable functions can be solved with a 3-level constraint in 1.65 min as opposed to 4.30 min with no constraints.

The previously observed degradation of running time with increase in the ratio of INPUT WEIGHT to OUTPUT WEIGHT in the cost function is reduced under constraints. Thus, for example, minimum interconnection networks were generated for the seventy-seven 3-variable functions in 3.65 min. Such a problem would require many hours with no constraints.

The capacity of the program to design multi-output networks was demonstrated by the generation of an 82-gate network realization of the seventy-seven 3-variable functions in 3.83 min. Analysis of the result and

[†]This work was supported by the National Science Foundation under Grant NSF-GK 1663 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

consideration of the first cost bound of 81 obtained by the algorithm indicates that 81 gates is the minimum possible network cost.

The algorithm has thus been demonstrated to be an effective tool for generating NAND networks both for logic-design automatization applications and for the study of the properties of NAND networks, e.g., diagnosability, usable fan-in, or regularity of interconnections, under constraints. The results of this effort, prior to the addition of constraints, are reported in full detail in Report R-382. [1] An article which includes the network constraint results is in preparation.

A comparison of the results of a recent paper on modular decomposition [2] with the results obtained by NAND decomposition followed by packaging into modules [3] indicated that the latter approach is faster and yields networks which are better than the "optimum" networks produced by the modular-decomposition algorithm. The difficulties with the modular-decomposition approach are caused principally by the selection of the library of available modules: The modules chosen are logically too complex and they are too numerous, hence the class of feasible solutions is extremely large. For the same reason, proving the optimality of the networks generated, which is done relatively easily in the case of NAND decomposition, is virtually impossible in modular decomposition. However, packaging an optimal NAND network into modules may destroy the optimality even in an environment where a direct relationship between the modules and identifiable subnetworks of NAND gates exists. Some preliminary results of the relative effectiveness of modular decomposition algorithms and NAND decomposition algorithms

followed by packaging and their range of applicability have been obtained. [4]

E. Davidson

14.2. Indistinguishability of Faults in Combinational Circuits

A technique for reducing the number of single faults to be considered for the diagnosis of a combinational circuit has been developed. The technique is based on the concept of indistinguishability classes of faults. [5]

The application of this technique to multiple faults is being studied currently. It is hoped that at least a partial solution can be found for the problem of which multiple faults will automatically be detected by a set of tests which was derived to detect all single faults.

D. Schertz

14.3. Generation of Fault Detection Tests for Combinational Circuits

A program has been written to generate tests which detect any logical fault in combinational circuits consisting of modules containing AND, NAND, OR, NOR, and NOT elements. As the program generates tests for a given fault, inputs to other logical elements are chosen to increase the chance of the test detecting several other faults. Thus a number of tests smaller than one test per fault is generated and less work is needed to reduce this set to the minimal covering set. A fault is labelled as indistinguishable if no test can be generated for it. Requirements for the

development of fault detection techniques in combinational circuits have been collected in a paper. [6]

Work is now being done to extend the procedure to diagnosis to the module level.

T. Powell

14.4. Fault Diagnosis to the Module Level

Work continues on the development of graphical and matrix notations which permit (1) a compact description of the logic functions performed by a module under normal and under faulty conditions, (2) the manipulation of these descriptions corresponding to the interconnections of the modules, and (3) the identification of faults in separate modules which cannot be distinguished without the addition of test point .

T. Gaddess

14.5. Selection of a Minimal Set of Diagnostic Tests

The development of a matrix procedure for the selection of a minimal set of diagnostic tests for single faults in multi-output combinational nets has been concluded and is reported in Report R-381. [7]

R. Isenhart

14.6. Design of Diagnosable Sequential Machines

A study has been made of the diagnosability of machines designed using Huffman's "one relay per row" canonical synthesis procedure. It has been shown that such machines have a uniform fault-behavior pattern which

can be easily detected. A simple diagnostic circuit can be attached which detects most faults (including all single faults) while the machine is in operation. Little or no fault analysis is required and fewer diagnostic test runs are needed.

Currently other synthesis procedures are being examined which, it is hoped, will be more economical and be applicable to a larger class of machines than the Huffman procedure.

J. Hayes

14.7. Fault Diagnosis in Cellular Arrays

A survey of cellular array literature was made. The problem of diagnosis of cellular cascades and cellular arrays has only been studied for combinational circuits, and even there the results obtained are of only preliminary nature. The diagnosis of sequential arrays and cascades has not been studied so far. Both problems are of interest because of the predicted spread of LSI technology in digital computer design.

For combinational arrays, necessary and sufficient conditions for diagnosis which are stronger than Kautz's conditions [8] have been found. A notation has been developed which allows (1) formulation of the diagnosis problem in a very natural way, and (2) definition of a matrix operation for building up the array from a single cell.

For the sequential cascades, an investigation has been started on the diagnosability of a cascade composed of individually diagnosable sequential elements. Conditions are being derived under which such a cascade is diagnosable from its boundaries.

S. Seth

14.8. Equivalence of Regular Expressions

Regular expressions have been used to describe the external behavior of sequential machines because they explicitly define all input sequences accepted by a given machine. Problems exist, however, in the manipulation of regular expressions. In particular, the problem of checking the equivalence of regular expressions has been studied and an equivalence-testing algorithm which is particularly suited to a class of very long regular expressions has been developed.[9,10]

J. Hayes

14.9. Ternary Logic for Asynchronous Sequential Networks

A study of speed-independent asynchronous circuits investigating the feasibility of using ternary logic and ternary-logic elements for the design of an arithmetic control comparable in function to the arithmetic control of ILLIAC II has been concluded.[11] The fundamental theory of ternary-logic algebra is presented, along with function representation and a discussion of minimization of ternary functions. A ternary design technique is also developed. A basic description of the ILLIAC-II arithmetic control is given for purposes of understanding design criteria and of establishing a motivation for using ternary logic. Finally the logic design of Basic Logic Blocks is discussed and problems and suggestions for future work are considered.

Ternary logic and ternary-logic design is being studied further. Topics being considered include: Rigorous formulation of speed-independent switching theory for ternary logic and for many-valued logics in general;

Universal ternary-logic elements; Design of combinational and sequential ternary-logic networks; and the development of a method for synthesizing hazard-free ternary circuits. Some preliminary results on the use of ternary memory devices and applications of ternary-logic control networks have been obtained. [12]

A. Wojcik

14.10. References

1. E. S. Davidson, "An Algorithm for NAND Decomposition of Combinational Switching Functions," C.S.L. Report R-382, May 1968.

2. P. R. Schneider and D. L. Dietmeyer, "An Algorithm for Synthesis of Multi-Output Combinational Logic," IEEE Trans. on Computers C-17, 117, (February 1968).

3. E. S. Davidson and G. Metze, "Comments on 'An Algorithm for Synthesis of Multiple-Output Combinational Logic'," IEEE Trans. on Computers C-17, in press, (December 1968).

4. E. S. Davidson and G. Metze, "Module Complexity and NAND Network Design Algorithms," to appear in Proc. Sixth Allerton Conference on Circuit and System Theory, October 1968.

5. D. R. Schertz and G. Metze, "On the Indistinguishability of Faults in Digital Systems," to appear in Proc. Sixth Allerton Conference on Circuit and System Theory, October 1968.

6. T. J. Powell, "Synthesis Requirements for Fault Detection," to appear in Proc. Sixth Allerton Conference on Circuit and System Theory, October 1968.

7. R. D. Isenhart, "Optimal Test Sets for the Diagnosis of Multiple Output Combinational Nets," C.S.L. Report R-381, May 1968.

8. W. H. Kautz, "Testing for Faults in Combinational Cellular Logic Arrays," Proc. Eighth Annual Symposium on Switching and Automata Theory, Austin, Texas, October 1967.

9. J. P. Hayes, "On the Equivalence of Regular Expressions," C.S.L. Report R-374, April 1968.

10. J. P. Hayes, "An Algorithm for Determining the Equivalence of Regular Expressions," to appear in Proc. Allerton Conference on Circuit and System Theory, October 1968.

11. A. Wojcik, "An Application of Ternary Logic to Logic Sequencing," C.S.L. Report R-388, July 1968.

12. A. Wojcik and G. Metze, "A Re-examination of Ternary Switching Theory," to appear in Proc. Sixth Annual Allerton Conf. on Circuit and System Theory, October 1968.

H. Von Foerster
R. T. Chien
D. E. Carroll
B. Carss
F. P. Preparata
S. R. Ray
E. J. Scott
P. Weston
H. Bielowski
K. Biss
J. Chow
P. Duran
G. Fariss
G. Goodman
C. Hartmann
J. Jansen
A. Kisylia
F. Loew
D. Lombardi
J. Omundson
P. Reynolds
J. Schultz
F. Stahl
S. Taylor
B. Wang
S. Wax

15.1. Research on R2 System

15.1.1. Introduction

The R2 system is a general retrieval system based on the "Rules of the Road" manual of the State of Illinois. The objectives of the R2 system program are twofold. First, the system is to provide us with a testing facility and data for the development and evolution of fundamental concepts in cognition. Second, it is to provide us with a general question-answering system with significantly broader scope than systems presently known.

During the previous period, a number of significant achievements have been made in the areas of relational structuring, parsing programs, context-modeling, and query analysis. The details of these results are summarized in the following sections.

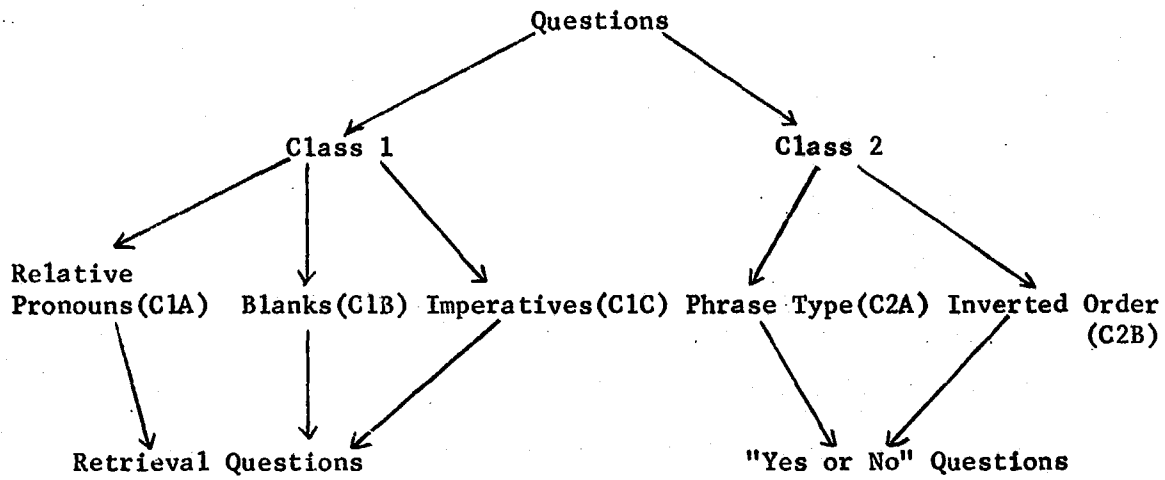
R. T. Chien

15.1.2. Query Classification

The key step in a question-answering system is to understand what the question means. When the question is understood, its meaning can be processed with a retrieval system, which will return an answer to the question. Our problem has been how to determine at least partially what the

†This work was supported by the Office of Education under Grant OE C-1-7-071213-4557 and by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

question means by normalization. It is found that we could normalize all questions in a step-wise manner as given by the following diagram:



First, we divided the questions into two classes. Class 1 questions would typically be answered by a sentence containing some retrieved information, while Class 2 questions would be answered by "yes, no, true, false." Class 2 sentences could be immediately recognized by the presence of such phrases (C2A) as "yes or no," "true or false," or by the presence of an inverted order, subject-predicate structure (C2B). The inverted order, subject-predicate structure, can be recognized in all cases by the presence of an auxiliary verb (must, can, is, does, etc.) at the beginning of the sentence or after a comma in the sentence, e.g.,

"On an interstate highway, can a car travel at 40 mph?"

"Must you have automobile insurance in order to drive?"

Class 1 consists of all other questions. It is characterized by

the presence of a relative pronoun (C1A), a blank (C1B), or an imperative (C1C). Class C1C is recognized by the presence of a word such as "list, give, state" at the beginning of a sentence or after a comma, e.g.,

"State the difference between a car and a motorcycle."

Phrases such as "would you please" in

"Would you please state the difference between a car and a motorcycle?"

are disregarded.

Class C1B sentences are "fill-in the blank type," e.g.,

"An octagonal-shaped sign means _____."

They are recognized by the presence of a blank.

Class C1A questions always contain a relative pronoun such as "what, when, where, who, whom, which, why, how, how much, how many," e.g.,

"What is the minimum speed for a car on an interstate highway?"

After making this initial breakdown into classes, (C1A), (C1B), and (C1C) are normalized so that the new form of the question starts with the relative pronoun (or interrogative word), or contains it in an initial prepositional phrase. Thus:

"If 4 cars arrive at a 4-way stop at the same instant, who has the right-of-way?"

becomes:

"Who has the right-of-way if four cars arrive at a 4-way stop at the same instant?"

while

"A no-passing zone is marked by what on the pavement?"

becomes:

"By what, on the pavement, is a no-passing zone marked?"

If a question is not Class 2, C1B, or C1C, then it must contain a relative pronoun. If there is only one relative pronoun, this pronoun must "ask the question," and not be in a subordinate position. When there are more than one relative pronoun, we normalize the sentence so that the initial relative pronoun "asks the question," i.e.,

"When the yellow light is on, and you have just entered the intersection, what should you do?"

becomes:

"What should you do, when the yellow light is on, and you have just entered the intersection?"

In this case, "when" gives information and doesn't ask for information as "what" does.

A program has been written to do this classification and normalization of questions automatically. The data consists of 195 questions directed towards the "Rules of the Road" as asked by 7 different people. All questions are classified as follows:

	Number	Percentage
Class I		70.8
Relative Pronouns	124	63.5
What	54	27.7
Why	4	2.1
Which	10	5.1
How How many How much	24	12.3
When	19	9.7
Where	2	1.0
Who For whom Whose	11	5.7
Blanks	14	7.3
Imperatives	7	3.6
Class II	50	29.2

At this point further analysis is applied to a subset of questions, namely, the subset of questions containing the relative pronoun "what." This subset seems attractive for two reasons. First, a relatively large percentage of the questions are of this type. Secondly, each of the relative pronouns can be expressed in terms as follows:

Original	Replacement
What	What
Why	For what reason
Which	What
How	In what manner
How many	What amount
How much	What amount
When	At what times
Where	In what places
For whom	For what person
Whose	What person's

In respect to the "Rules of the Road," these correspondences are much more specific, e.g., "where" might correspond to "on what road."

The "What" questions fall into four sub-classes from which to retrieve meaning.

1. Questions of the form: "What N V S"

Example 1A:

"What markings are used for the center line of a four-lane highway?"

This sentence would be deciphered as:

N = "markings"

V = "are used"

S = "for the center line of a four-lane highway"

The set of all things equal (modulo V) to S would then be found, and the intersection of this set with N would give the answer. This set might include: division, double yellow line, median, etc. Since double yellow line is a marking, this would be the answer. It is not claimed that the intersection would be singular, and the number of elements in the intersection would be a measure of the ambiguity of the question.

Example 1B:

"What signs are associated with a no-passing zone?"

Interpretation:

N = "signs"

V = "are associated with"

S = "a no-passing zone."

The set might include: danger, SIGN: DO NOT PASS, slow down, yellow line next to dashed white line, etc.

Answer: SIGN: DO NOT PASS.

2. Questions of the form: "What V_1 N V_2 S"

Example 2:

"What do you do when you approach a red light?"

This question is among the last four questions that were added in anticipation of questions that could be asked, but of which we had no examples. For interpretation:

$V = V_1 \cup V_2$

and this sentence is of the form, "What V N S." Then the set might include: stop, decelerate, shift gears, etc.

Answer: "Stop, shift gears."

Since when NOT S occurs, you also shift gears but you don't stop, the best answer, stop, could be arrived at. This method of eliminating ambiguity probably would not always work, but it looks promising as a first approximation.

3. Questions of the form: "What be (V) S"

Example 3A:

"What is the meaning of a small white rectangular sign marked with the letters JCT?"

This sentence would be deciphered as:

be = "is"

(V) = "meaning" (V) = ϕ means: definition

S = "of a small white rectangular sign marked with the letters JCT."

The set of all things equal (modulo V) to S would then be found. In the computer dictionary we would assume that "meaning" is the same as "equals," giving the set: crossroad, danger, etc.

Example 3B:

"What is a motor-driven cycle?"

The set equal to "motor-driven cycle" would then consist of: "a motor vehicle with two wheels with less than 150 c.c. engine displacement."

4. Questions of the form: "What AV N (V) S"

Example 4:

"What must a driver do before passing through a green light?"

The sentence would be deciphered as:

AV = "must"

N = "a driver"

V = "do"

S = "before passing through a green light."

Then S would imply (modulo V) a set of things, whose intersection with N

would give an answer. The set might include: decelerate, slow down, get into the proper lane, etc.

In this case the answer would be: "slow down, get into the proper lane."

In the future, we plan to look at what we call "S." Current thoughts are towards feeding "S" into a parser, or looking at keywords contained in "S." Since how we get meaning out of "S" appears to be closely related to how the "Rules of the Road" is structured into memory, we plan to look at this also.

P. Reynolds
C. Hartmann
R. T. Chien

15.1.3. Syntactic Analysis

This work is a continued effort to find fast and efficient methods of assigning a structure to natural English sentences, especially those sentences found in the Illinois Driver's Manual--"Rules of the Road," so that when stored in a computer, the machine will be able to answer questions on "Rules of the Road."

In the Sept. '67 to Feb. '68 Progress Report, we outlined a method for normalizing natural English text. Let us briefly review the method.

The set of sentences of a document are divided into two categories predications and non-predications. Predications are sentences of the form "A is in some way a function of B" and non-predications are sentences which do not have this structure. A particular function in a predication can be expressed in many ways. In order to normalize the text we replace words that

express the function by the function itself. Thus in the sentence "A car is an automobile" the word "is" expresses the function equality. The normalized sentence then becomes "A car equals an automobile."

We tried to apply the above method to the sentences of "Rules of the Road," hoping that those sentences which were non-predications would not contain much, if any, information, and those which were predications would be easy to normalize. However, we found that only a few sentences from "Rules of the Road" could be normalized in this way and that the normalization was very involved due to the nature of the sentences. For these reasons it was not considered worthwhile to attempt to program this method of sentence normalization and we began looking at dependency analysis and immediate constituent analysis.

The basis of immediate constituent analysis is that adjacent sub-strings of a sentence are syntactically related. The relation is defined by the rules of the immediate constituent grammar used in the analysis. These rules are of the form $A_j + N \rightarrow NP'$ which indicates that an adjective and a noun appearing next to each other in a sentence can be combined to form an NP'. For example, let us analyze the sentence "The old man sat on the bench" assuming we have the rules:

1. $V + PP \rightarrow VP$
2. $P + NP' \rightarrow PP$
3. $A_j + N \rightarrow NP'$
4. $T + NP' \rightarrow NP$
5. $NP + VP \rightarrow S$

Putting parentheses around the combined terms, we would get the following parsing:

(The (old man)) (sat (on (the bench))).

This analysis can be represented by the tree of Fig. 15.1, although we then lose the information about the order in which the rules are applied.

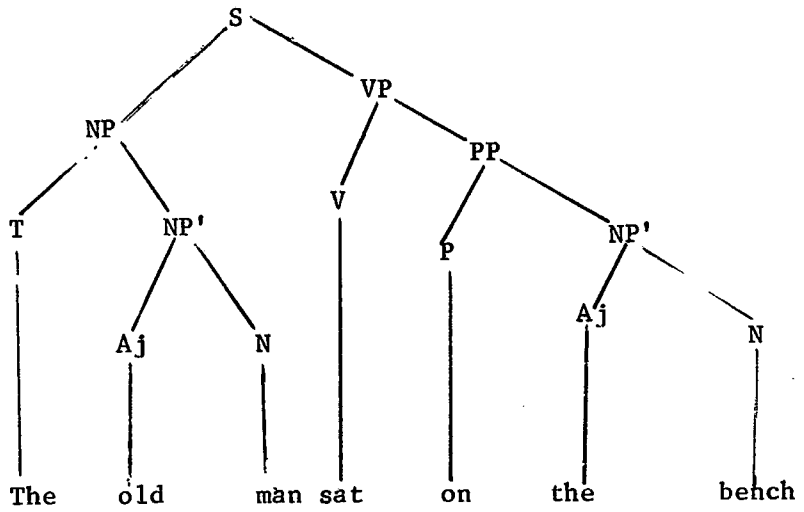


Fig. 15.1.

In our immediate constituent analysis program, our method is quite similar to that of John Cocke [1] of IBM Research. The basic plan of this approach is to build all possible two-occurrence phrases, then all possible three-occurrence phrases, etc. The syntactical structures of an n-occurrence system are the k-occurrence phrases.

The input to the program is the individual words of a sentence. First, all the words of the input sentence are looked up in a dictionary to assign the words all possible parts of speech. Next, all consecutive

pairs of 1-occurrences of words are examined with all possible grammatical interpretations. For example, let us parse the sentence "The old man sat on the bench."

Form	Grammar Code Symbol
The	T
old	A
man	N
sat	V
on	PP
the	T
bench	N

In this example the first two 1-occurrences, T-A, are examined and are tested for agreement. The test is made by table look-up. The table that Cocke uses is a list of ordered triples of the form $A + N \rightarrow NP'$. If the pair shows agreement, it becomes a 2-occurrence phrase and is stored. If not, the second word is checked with the third A-N. After all 2-occurrence phrases are stored, the program is checked for 3-occurrence phrases. The program performs this by checking the 2-occurrence phrases with the consecutive 1-occurrence phrases and vice versa. The program continues on the length of the phrases to be constructed until phrases as long as the sentence have been constructed. At each occurrence length, say X, it looks at all occurrence phrases of length n, n=1,2,...,X-1, immediately followed by X-n. For our example the output of this program would be:

Item	Starting	Location from	Length	Grammar Code	Constituent Items
1	1	The	1	T	
2	2	old	1	A	
3	3	man	1	N	
4	4	sat	1	V	
5	5	on	1	P	
6	6	the	1	N	
7	7	bench	1	N	
8	2	old man	2	NP'	2,3
9	6	the bench	2	NP'	6,7
10	5	on the bench	3	PP'	5,9
11	1	the old man	3	NP'	1,8
12	4	sat on the bench	4	VP	4,10
13	1	the ... bench	7	S	11,12.

The main differences between our program and Cocke's are: a) our program is written in CSL6 language while Cocke's is written in Fortran and b) our program will give only 1 parsing. We eliminate the possibility of several parsings by allowing each constituent to be a member of only one larger constituent. That is, once a particular constituent is integrated into a larger constituent, it is no longer a candidate to appear in any other constituent.

The reason why we chose Cocke's program is because it is so much faster than any other immediate constituent analyzer. We felt that we can

even improve upon the time by not caring about ambiguities.

A dependency analysis consists of finding a tree which represents the structure of the sentence being parsed. The main verb of the sentence is taken as the head node of the tree with other words in the sentence appearing at the other nodes of the tree. The connections between the nodes indicate dependency where word A depends on word B, or B governs A, if A modifies or complements the meaning of B. For example, the sentence "The old man sat on the bench" would be represented by the following tree:

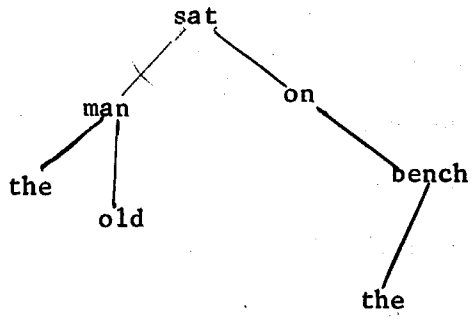


Fig. 15.2.

In this example "old" and "the" depend on "man," while "man" depends on "sat," etc.

Let us look more closely at the tree of Fig. 15.2. If we drop a projection line from each node to a base line, we see that no projection line crosses a dependency link (see Fig. 15.3).

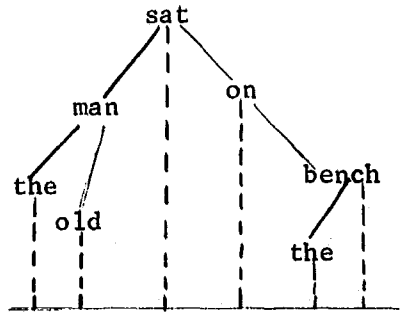


Fig. 15.3.

This property is not a universal property of English sentences, however, the assumption that all sentences satisfy the property of projectivity is like the assumption that a sentence is made up of immediate constituents.

Hays [2] has written a parsing program using a dependency grammar. The program works on the premise that two words will not be connected in a dependency tree if this connection will lead to a non-projective tree.

The first step in the program is to find a grammar code symbol for all words in the input sentence by dictionary look-up. Then using the string of grammar code symbols the program tests all pairs of symbols for precedence where precedence is defined in the following way:

X precedes Y if and only if

1. X is to the left of Y in the sentence
2. Any words between X and Y depend on either X or Y (This insures projectivity.)
3. Neither X depends on Y or Y depends on X (This insures that there will be no circular connections.)
4. Either X or Y or both are independent.

The program then makes the following computation on all pairs of words to find if X depends on Y.

X depends on Y if and only if

1. X is independent
2. X precedes Y or Y precedes X
3. There is an entry in the dependency table which states that the grammar code symbol of X can depend on the grammar code symbol.

When a dependency connection is made, the grammar code symbols for both X and Y are changed to show the properties of the pair. In this way Hays was able to obtain an immediate constituent analysis along with the dependency analysis. The final output of the program is all possible dependency parsings for the input sentence along with the associated immediate constituent analysis.

Simmons [3] has also written a program called the Pattern-Learning-Parser (PLP) which uses a dependency grammar. All of the grammar rules which this program uses are learned from a hand-parsed text.

To be more specific, the words in the hand-parsed sentences are each given a number. This number indicates the level at which the word occurs in the dependency tree of the hand-parsed sentence. For example, the words of the sentence "The old man sat on the bench" would be assigned the numbers which appear below the words (note that the words on the object side of the verb are positive). The program then generates, for each word in the hand-parsed text, a word class of the form a/b/c and a set of sentence-pattern

rules which indicate the structure of the dependency tree of the hand-parsed sentence. The word class a/b/c that is associated with a particular word indicates that the word occurred at level a in a tree, that the word appearing ahead of it was at level b, and that the word that followed it appeared at level c.

When a sentence is given to the computer to parse, the word classes for each word in the input string are looked up. By successively applying logical addition and logical subtraction to the word classes a prediction is made about the depth at which each word of the input string might appear. If there is an ambiguity, the sentence-pattern rules are applied.

This program gives multiple parsings for an input sentence, but Simmons only uses the first possible analysis.

It has been proven by Gaifman [4] that dependency grammars and immediate constituent grammars are weakly equivalent. That is, they generate the same set of sentences from the same initial vocabulary or, analytically, they classify the same string of sentences. However, it can be shown that dependency grammars and immediate grammars are not strongly equivalent. That is, there does not exist an isomorphism between the structural diagram each associates with a given sentence. Since an isomorphism does not exist, a sentence which may be ambiguous (have several structural diagrams associated with it) in an immediate analysis may have but one dependency tree, and vice versa. Therefore, dependency analysis and immediate constituent analysis each capture something of grammar which the other misses. It is for this reason Hays chose to use both dependency analysis and immediate

constituent analysis.

Part of our project is to find out if it is really necessary to use both types of analysis. We are not completely convinced that, in the R2 system, two types of structural analyses will give much more information than one type of analysis. It is to test this premise that we are developing both an immediate constituent parsing algorithm and a dependency-parsing algorithm.

The dependency-parsing program is running at this time. Our program is fast and discovers, for any input sentence, the most probable parsing. The program works in the following way.

The words of the input sentence are looked up in a dictionary to discover their syntactic word classes. Using the string of syntactic word classes, we search for the governor of a particular word, checking the words first at distance one from the given word, then at distance two, etc. A word and its candidate for governor are checked in a table of possible dependency connections to find if the syntactic class of the candidate can govern the syntactic class of the particular word whose governor we are trying to find. Once a governor is found for a word we make the connection between dependent and governor and go to the next word.

We have chosen this way to parse over the approaches of Hays and Simmons for several reasons. Both Hays' program and Simmons' program gives multiple parsings. Simmons arbitrarily picks the first parsing as the correct one. By searching for the most probable governor, we eliminate

questions of ambiguity but still get a "good" parsing quickly.

At this time our program can deal only with simple sentences. However, we are working on subroutines to deal with compound and complex sentences.

K. Biss
J. Schultz
R. T. Chien

15.1.4. Context Modeling

In trying to realize a model that could be used to logically structure the "Rules of the Road" in a data base, it was first necessary to make some preliminary judgments concerning the major divisions to be handled. It was found the entire text could be divided into 5 categories that would span all conceivable topics. These 5 categories are listed below:

1. legal requirements
2. motor vehicles
3. traffic laws
4. accidents
5. unsafe driving practices and emergency action

These categories are obviously not mutually exclusive, but each has distinct constituents that make it a definite advantage to try to deal with them separately.

In an attempt to derive a successful model for arranging information concerning traffic laws, to begin with, it was necessary to make a further classification. Traffic laws can be subdivided into the following topics that are necessary (although possibly not sufficient) for an adequate

coverage.

1. right of way
2. passing
3. lane usage
4. turns (& signals)
5. speed restrictions
6. stopping
7. parking
8. trucks, buses, trailers, taxis, etc.

A further inspection of these subtopics reveals that each can be broken down into two basic components concerning traffic law: 1) an explicit component, that can be uniquely defined by a highway sign, a traffic signal and/or a pavement marking; and 2) an implicit component that is known a posteriori. For example, there are specific highway signs and signals that regulate turning and indicate exactly where and when it is legal to make a turn. However, to turn safely and obey the traffic laws, it is also necessary to make up your mind before you get to the turning point, move into the proper lane as soon as possible and to slow down to a "reasonable" turning speed. These last three prerequisites for a safe and legal turn comprise the implicit component of the subcategory "turns (& signals)." Therefore, it is apparent that if some systematic structure for modeling traffic laws is to be devised, it must be capable of handling both of these components successfully.

Some work has already been done on modeling the explicit component.

This was a relatively simple task since in this category we have a physical basis for establishing the various relationships between the entities involved. As an example, a model for arranging information concerning highway signs was devised and is described below.

It was first necessary to develop an attribute-value property list for each sign. It was found that each sign analyzed could be completely described using five-attribute-value pairs. These pairs are listed below for a STOP sign.

Stop

(shape: octagon)
(color: red with white letters or yellow with black letters)
(location: intersections)
(purpose: regulate traffic)
(driver reaction: to bring vehicle to a complete stop before entering a marked crosswalk or crossing a marked stop line. After stopping he must not start again until he has yielded the right of way to pedestrians and to approaching traffic on the through highway, and until a safe interval occurs)

Any information concerning an attribute of a particular sign can be modeled using an attribute-value property list similar to the one shown above. For example, the above list contains all the necessary information for answering a question of the type: "What is the shape of a STOP sign?". However other types of questions concerning signs must also be evaluated. If the question does not refer to a specific sign, but is in regard to the general nature of signs, the model must be capable of keying on certain words and making the appropriate information available. This means that it is necessary to interrelate all the information in such a way as to make all the

necessary information available regardless of the point where the data array is entered.

The following diagram Fig. 15.4 illustrates how this information can be interrelated.

It is easy to see how a model of this type could be extended to include traffic signals, pavement markings and, in fact, any physical aspect of driving that can be described by attribute-value pairs. However, when it is attempted to extend this model to include the implicit component, it fails. At present, a scheme based on cause-effect relationships is being tried to see if it would permit a systematic structuring of the implicit component. Once a successful scheme is devised, it will be necessary to try to effect a compatible intersection between the different models and thereby lead to a successful overall model.

R. T. Chien
D. Lombardi
F. P. Preparata

15.1.5. Concept Processing

Initial processing has been completed on the data base for "Rules of the Road" with respect to the maximal phrase relational strategy. The objective is to find a general strategy technique with a high relevancy rate for retrieving statements from the data base, given a question that can be answered by using the statements from the data base.

A strategic relevancy rate has been defined as N/(A+E); where A is the number of statements in the data base that are relevant; N is the number of relevant statements that the given strategy technique actually encountered,

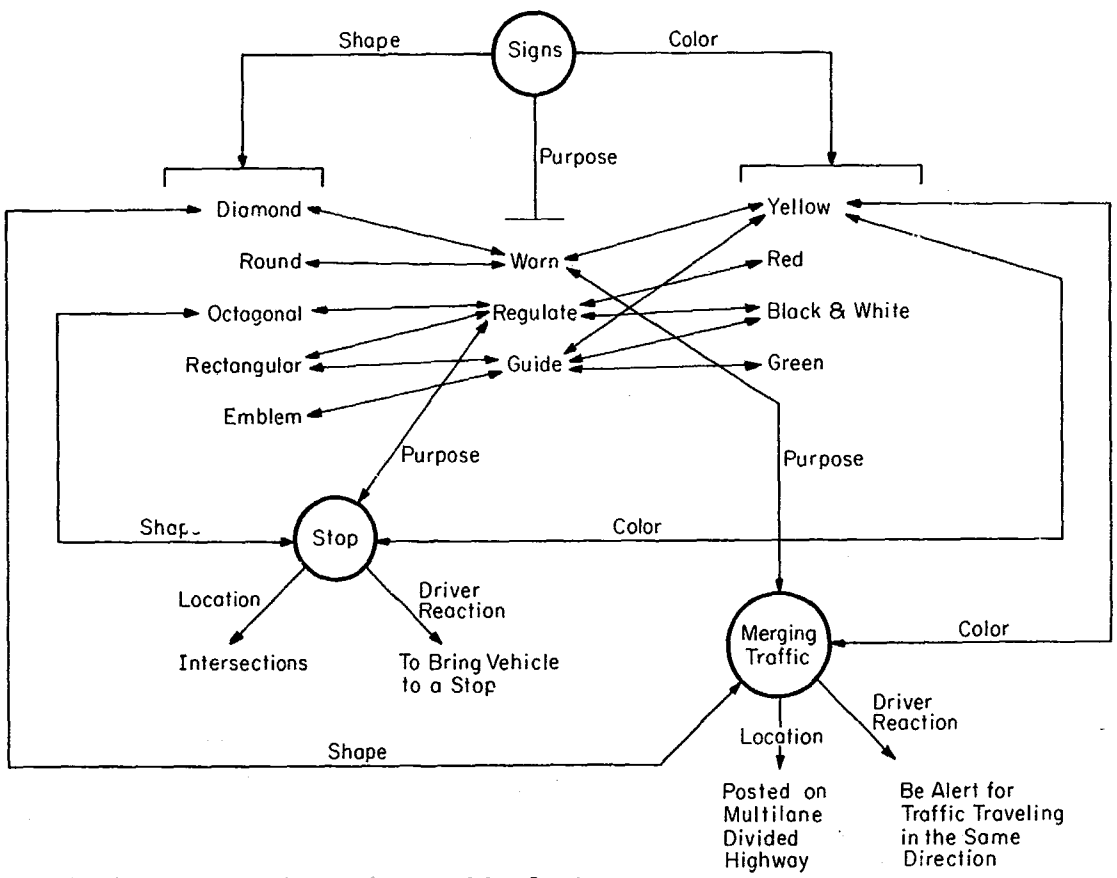


Fig. 15.4. A model of the world of signs.

note $N \leq A$; E is the number of statements that the given strategy encountered, but that are not relevant.

The maximal phrase relational strategy consists of a dictionary, a set of statements, and a strategy.

Each entry in the dictionary contains a maximal phrase and a set of pointers that refer to statements.

Maximal phrases are arrived at in the following manner:

a. Sentences of the data base are numbered consecutively.

b. A WIS Index is processed on the sentences from a. WIS means words in sentence, its name is derived from KWIC where we consider all words not just key words and the context is a sentence.

c. Consecutive entries of the Index are compared to find the longest string of words that match. Thus, if STOP AT A STOP SIGN...239
and STOP AT A RED LIGHT...646
were consecutive entries, then STOP AT A 239, 646
would be recorded.

d. Remove the prefixes from the output of C. If one entry is the beginning part of another entry, and if these two entries have any numbers in common, then the numbers that are common to both are removed from the former. If an entry results with no numbers, it is deleted, that is:

Stop 239, 362, 424, 749

Stop at 239, 646

Stop at a 239, 646, 932

Stand in 136, 426

would result as

Stop 362, 424, 749

Stop at a 932

e. The output from d is reverse sorted. That is, if an entry were

Stop at a

it would be sorted as if it were spelled

a ta pots.

This results in a list of entries that have similar endings like:

go at a

stop at a

halt at a.

f. The suffixes from the output of e are removed. This is an analogous operation to removing the prefixes in d.

Thus if the book 269, 348

and of the book 269, 348, 729

were entries, then

of the book 269, 348, 729

would be recorded.

g. The output of f is re-sorted in normal alphabetical order. There are a number of strategies that are being tested for relevancy rate. Given a question:

a. Find the maximal phrases of the question and retrieve the indicated statement.

b. Find the maximal phrases of the question and retrieve all those statements that occur with at least two pointers, three pointers, etc.

c. Retrieve statements by a, or b, and consider the maximal phrases of these statements and retrieve those indicated statements.

d. Same as c, only retrieve statements, with technique of double occurrence of b.

e. Find the maximal phrases of the question and for each pointer n also consider pointers n-1 and n+1 (the statements have same clustering!).

f. Same as e, using double occurrence of pointers as retrieval criteria.

F. Stahl
J. Jansen
R. T. Chien

15.2. Cylinders: A Data-Structure Concept Based on a Novel Use of Rings

A form of linked data structure called the "Halo" has been described. [5] A continuation of this effort has resulted in another and more versatile structural concept which we have called the "Cylinder."

15.2.1. Introduction

With the growth of non-numerical computer applications, and the attendant necessity of dealing with entities lacking the orderly internal relationships characteristic of matrices, interest in linked data structure has grown, with the key developments coming out of artificial-intelligence research where indeed the data-manipulation problems are the most severe.

Structuring of data with the computer store is a factor, implicit or explicit, in the design of any computer program. The point of tangency between data structure and program design lies in the necessity of imposing relationships upon data independently of their values, in order simply to express an algorithm or computing procedure upon them. This is particularly clearly seen in, for example, the case of matrix multiplication, where row and column relationships must be imposed in an appropriate way upon a set of stored numbers in order for the required procedure to be definable as a program. As a general principle, it may be said that, in the context of computation, data structure is dependent upon the processes to be performed, and is less an intrinsic quality of the data themselves.

In numerical applications it is ordinarily sufficient to structure data in matrix form provided that arrays with various numbers of dimensions can be handled, and compilers oriented toward numerical problems, such as FORTRAN and its successors, contain simple and direct means, including subscripting of variables and dimension declarations of arrays, for expressing the necessary relationships. What sets this type of structuring apart from linked structures is the fact that it is possible in an array to compute, in a reasonably simple manner, the location in the computer store--or address--which holds a given datum, on the basis of its description as an array element, i.e., the values of its subscripts and a "base address" corresponding to the variable itself. Indeed, even at the hardware level, current computers universally include means for dealing directly with one-dimensional structures through the use of address-modification registers.

When an algorithm effectively precludes the use of arrays, some form of linked structure is the only presently available alternative, since content-addressed stores--which are the other presently interesting possibility--are not now included in ordinary equipment and may never be. The difficulties which lead to abandonment of arrays are encountered when the details of the required structure are either; (2) highly irregular, leading to great quantities of wasted storage due to empty positions in an array, or (b) highly variable, leading to massive shuffling around of storage contents when array parameters are redefined. An important subcase of the latter is that in which structural requirements are substantially dependent upon data values, as happens in most instances of artificial intelligence problems. The automatic parsing of sentences can be taken as an illustration. The entire object of such an algorithm is the generation of structure upon the incoming data. In this case, the structure represents the syntactic relationships to be found in the input sentences, which are initially given simply as strings of symbols.

Because our work is primarily in the artificial intelligence area, the need of a good linked-structure system has been recognized as an important consideration. Two of the older and well-established linked-list languages, IPL-V [6] and LISP [7] were not felt to be entirely adequate, due mainly to the restricted class of linkage structures realizable, i.e., in both cases only unidirectional linking is possible and no reentrant linkages are allowed, with the result that from an arbitrary internal element in a linked structure only a portion of the remaining structure

is accessible. While for many purposes this objection is not actually as serious as it sounds, there do arise many cases in which complete availability of a structure from any element in it is a decided advantage. There have been attempts to provide more or less direct reversibility of linkage paths, e.g., Threaded Lists, [8] and the SLIP system. [9] The latter tends to impose heavy burdens upon memory space and the former involve some potentially severe problems due to untraceable connections when storage cells are "erased."

More recently, the usefulness of ring structures has been demonstrated, [10] and they offer several advantages over the earlier approaches. The principal ones are (a) as the name suggests, the basic structures are closed upon themselves, so that complete availability of structure is implicit in the design without a great increase in storage space, and (b) as they have been implemented, ring structures usefully combine linking and array structuring and in so doing make possible a fundamentally more efficient use of the computer. A further, most significant advance has been the invention of general-purpose, low-level, linked-list programming languages, the most widely distributed one being "L6". [11] In sharp contrast with earlier approaches, these are not tied to any particular linkage scheme, but are sufficiently general in concept and economical in the programming means they provide, to enable the user to realize whatever form of linked structure he may desire, using a fairly natural and compact programming notation.

Since December, 1967, L6 has been available at CSL on the CDC 1604

computer and a cylinder system has been programmed using it. This system, which we call CYL6, is now in active use as a tool in other programming work.

15.2.2. Ring Structures

In the terminology of linked lists, computer storage is dividable into blocks and fields, in addition to the common word unit. A block is a group of consecutive computer words, i.e., having consecutive addresses, while a field comprises an arbitrarily defined region internal to a block, having limits fixed with respect to the block boundaries but not necessarily coinciding with any word boundaries. A common schematic representation for these relationships is shown in Fig. 15.5. A linked structure may be created by using some fields in each block to hold the addresses of other blocks, i.e., to contain pointers. This is often represented in the way shown in Fig. 15.6.

As a list structure, a simple ring, which is the basic linked structure in the Cylinder system, is nothing more than a chain of pointers which closes upon itself. This can be represented as in Fig. 15.7. Although the circularity of this type of structure could apparently lead a program to step endlessly from link to link in the course of a data search, (and for this type of reason all reentrant structures were avoided in early linked-list systems), it is in fact a simple matter to prevent such a difficulty, when the ring is the only closed linkage pattern allowed.

This is typically done by incorporating a specially designated and identifiable cell in each ring to serve as a reference point in searching and building functions. A similar device was used in our earlier Halo

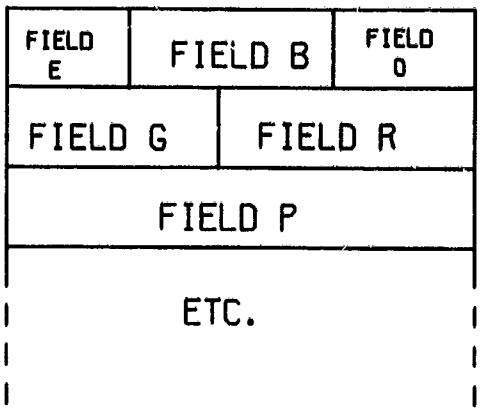


Fig. 15.5. A block is a group of consecutive computer words, within which areas called fields may be arbitrarily defined and labeled.

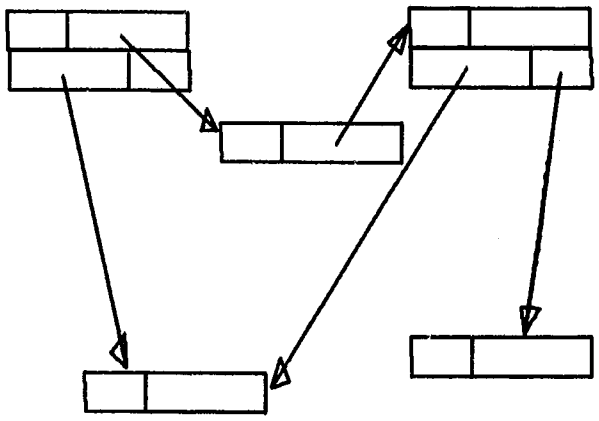


Fig. 15.6. A schematic convention for linked structures, using arrows to connect the blocks which are pointed to by stored links.

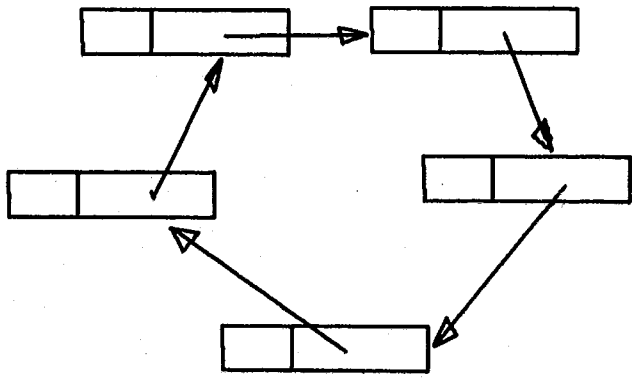


Fig. 15.7. A simple ring is a closed chain of pointers.

(OVERLEAF BLANK)

design, [12] but in the Cylinder system, emphasis is placed on eliminating all specially reserved storage for system functions--and most other restrictive system conventions--and the Cylinder ring is an entirely homogeneous structure. To insure termination of all ring-searching operations, the subroutines which execute them have been designed to automatically stop with a "fail" indication in the event that the starting point of the search is reencountered. This approach reduces storage overhead requirements and simplifies system conventions, simultaneously giving the user maximum latitude.

Given the choice of rings as the basic linkage pattern, which insures complete availability of each point in a structure from any other point in a minimally-complicated set of system conventions, there still remains a wide field of choice in the exact manner in which the rings are to be used in representing data structure. Common to all methods is the use of rings to "thread" blocks of storage which may contain data as well as ring pointers, linking the blocks together somewhat as if they were beads on a string. One direct application of this basic scheme is in representing a classification of data. See Fig. 15.8. Using individual blocks to hold descriptive information for the data, the representation is constructed by linking together on rings those data blocks which belong to each subclass to be represented. Because room can be made for more than one ring to thread a block, thereby representing membership of a datum in several classes simultaneously, there is no difficulty in setting up arbitrary hierarchical or overlapping classifications upon a set of data.

Using ring structures in this way requires little or no more storage than conventional lists do, while in contrast with conventional lists, rings allow full flexibility in the manner in which data may be searched, i.e., with equal ease a case (ring) may be searched for the data (blocks) it contains, or a datum may be searched for the classes which contain it.

15.2.3. Cylinders

Further discussion of the use of rings in creating data structures will be limited to the context of the cylinder conventions. These are:

(a) The only allowed linked structure is the ring, i.e., a non-branching, self-closed chain of pointers.

(b) A fixed number of fields, eight in the case of CYL6, are defined upon the first words of all blocks, and all ring pointers within the system must be contained in just these fields. However, there are no system-imposed restrictions on the use of fields which are not needed in a given program for building rings, i.e., if only a few levels are needed, short blocks may be used, or data may be stored at the unused levels, wherever in the block they may occur. Program context is the determining factor. In CYL6 the eight ring-pointer fields are identified by the letters "I" through "P".

(c) All of the pointers constituting any one ring lie in the same field within their respective blocks, i.e., if one pointer in a ring is contained in a "K" field, all pointers in that ring necessarily lie in "K" fields. It is convenient to speak of the allowed pointer fields as "levels" at which rings may be constructed and to speak of a ring at a certain level, e.g., in CYL6 a particular ring might be at the "K" level or "I" level, etc.

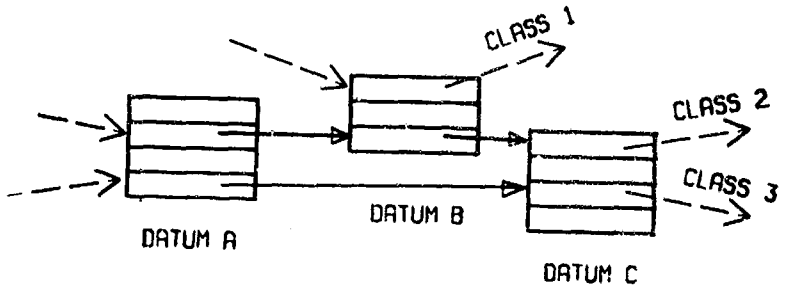


Fig. 15.8. Threading of blocks by multiple rings to represent a classification upon a set of data.

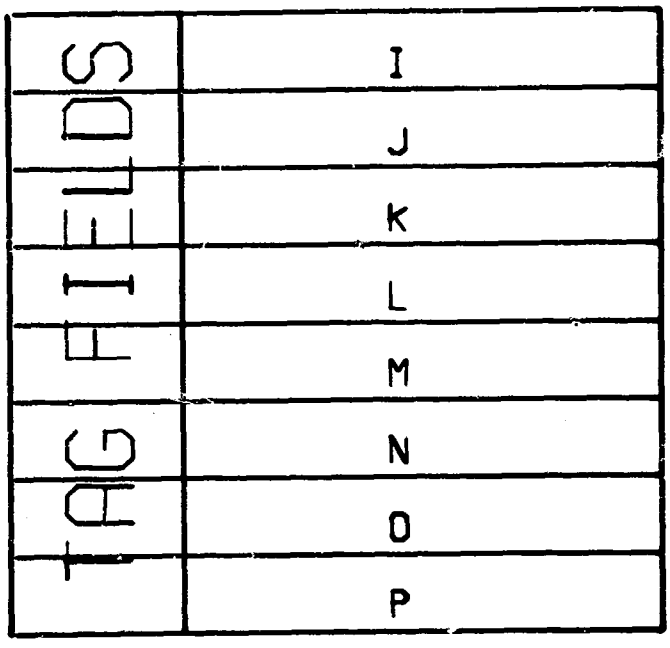


Fig. 15.9. Conceptual scheme of pointer levels as defined in the CYL6 system. This is not the actual arrangement in storage.

	I	Ø	J
U	K	V	L
W	M	X	N
Y	O	Z	P

Fig. 15.10. Actual storage arrangement in CYL6, using the first four words in a block. Field O is the tag field shared by levels I and J. Fields U through Z are the tag fields for levels K through P. The first six bits of the first word are reserved for the CSL7 storage allocator.

(OVERLEAF BLANK)

See Fig. 15.9.

(d) Associated with each pointer field there is also a data field of limited size, called a tag field. All data search functions in the CYL6 system are sensitive only to the contents of the tag field associated with the level being searched. All tag contents are completely under control of the user and are primarily intended to identify, for search purposes, functionally distinct types of storage blocks within the context of an individual program. In many applications tags can provide all of the identifying information needed to govern every desired mode of search in a structure. In CYL6, the tags have a capacity of nine bits, which may be divided into independent three and six bit subfields or may be taken altogether for search purposes. It is convenient to describe the contents of a CYL6 tag field with an octal digit followed by a BCD character, e.g., "5A," "1/."

A peculiarity of the CYL6 system is that while there are eight ring levels there are only seven tag fields, the I and J levels having a single tag field in common. This is due to interaction with space reserved in the first word of each block by the storage allocator of CSL7, [13] a dialect of L6 in which CYL6 is written. See Fig. 15.10.

The four conventions listed above are the only restrictions which are required by cylinders. They define a system broad enough in scope to include most other ring structures, and, in the case of CYL6, they are able to coexist with any other type of structure possible in L6, provided that the CYL6 routines are used only on rings of the cylinder type.

To gain further understanding of the structures that are available through the cylinder system it will be helpful to use an appropriate notation, such as the one which grew out of the CYL6 development effort. Figures 15.11 through 15.15 help to demonstrate this notation and its interpretation in terms of the block-and-pointer notation of the earlier figures. In this scheme the chain of pointers constituting a ring is represented simply by an arrow, the direction of which indicates the sequence of the indicated data elements along the ring. Since it is understood that all pointer chains are closed upon themselves, this fact may be kept in mind making it unnecessary to draw the rings as closed figures. When a single block is threaded by rings at several levels it is natural in the cylinder context to treat such a block as a cross-linking between rings and this is the way it is drawn, as in Fig. 15.11.

Although it is in fact generally the case that a structure consisting simply of cross-linked rings will tend to be unwieldly in any notation, some order can be extracted from the chaos by limiting one's attention to a single pair of rings at a time. In this case, the possibilities are quite limited, a block is threaded by none, one, or both of the rings, and in the latter two cases may or may not be threaded by other rings as well (Fig. 15.12).

The basic cylinder is formed from a pair of rings and employs cross-linking between them in a particular way. The framework of the scheme is pictured in Fig. 15.13, where the triangles are used to indicate the positions on each ring of the special cross-links which we have called

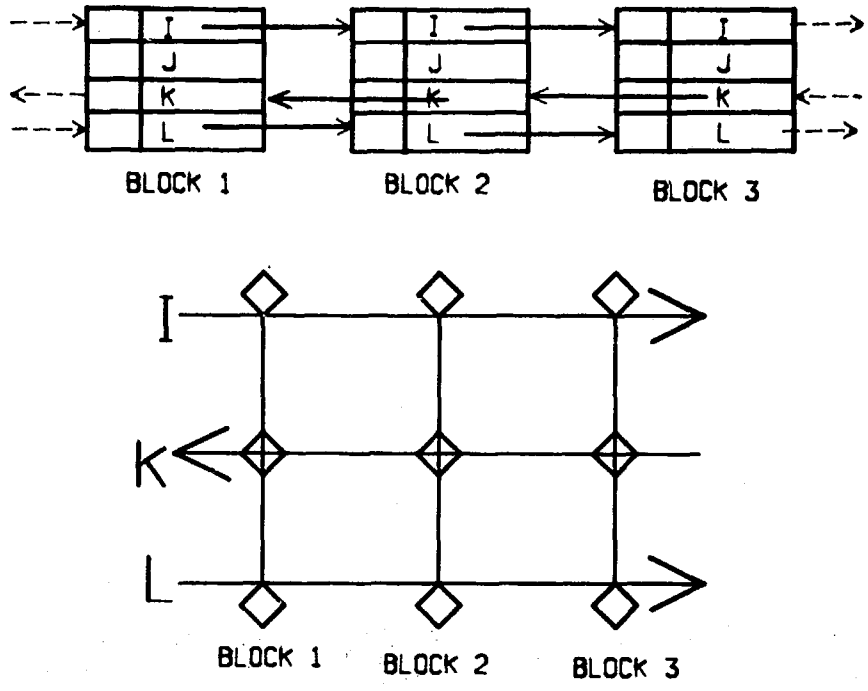


Fig. 15.11. A schematic convention for cylinder structures. Above is the block and pointer equivalent of the lower diagram. The long arrows are understood to signify closed rings of pointers running in the sequence indicated by the arrow direction.

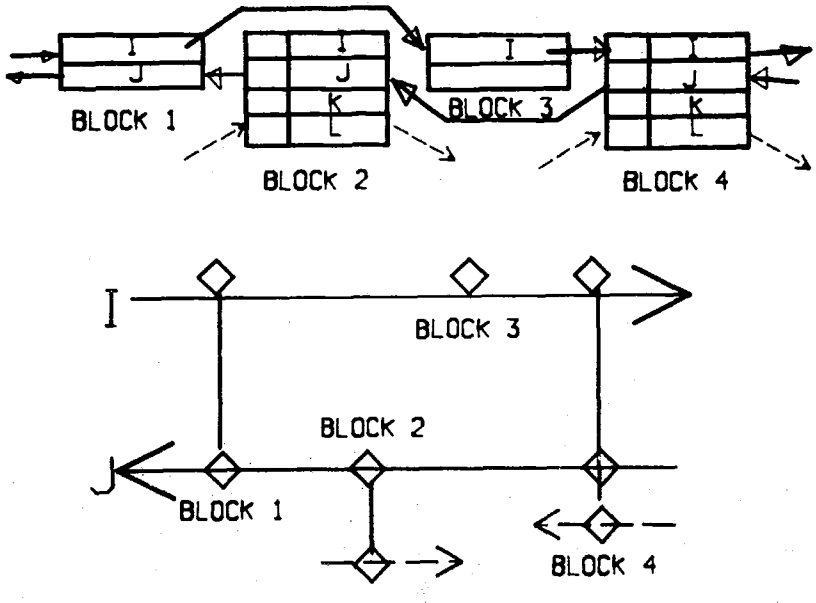


Fig. 15.12. The basic linkages possible when a particular pair of rings is taken separately. The particular choice of levels here is unimportant.

(OVERLEAF BLANK)

"seams." The only defining property of the seam substructure is that the blocks which form the seams lie either in exactly identical or more usually in exactly reverse sequences on the two rings. In the figure, the arrow directions shown indicate this ordering. The existence of such seams is not a mandatory feature in the CYL6 system. Their insertion, removal, and selective tagging is left entirely up to the user's discretion.

The usefulness of having a sequence of identifiable seam cross-links with the indicated ordering lies in the fact that between each pair of seams a type of subring is created which has the important characteristic of self-closure as does the main ring but has several added features of some usefulness as well. As Fig. 15.13 shows, each subring contains four segments, consisting of the two sequences of ring links lying between the two bounding seams at the lower and upper levels respectively, and of the two seams themselves, which constitute bilateral connections between the upper and lower ring segments. One natural way of using this ring and subring configuration to represent data structure is to assign a datum to each seam and to represent binary relations between data by other cross-links which carry tags identifying the relations they represent and are threaded by the lower ring segment of one subring, whose seam represents the value of the first argument of the relation and by the upper ring segment of the subring for the datum at the second argument position. Such a representation is shown in Fig. 15.13.

This representation for binary relations is highly compact as a linked structure, and is quite complete in representational power in view

of the facts that both the forward and inverse relation are simultaneously represented, and that the main rings constitute a master file from which all the data or all the links representing any particular relation may be selectively retrieved, using a tag-controlled data search. The storage requirement in terms of number of necessary links compares quite favorably in this case with that for IPL-V, provided that all the above information is to be represented. In the case of IPL-V there is a basic requirement of two links per datum, plus eight links to represent both a binary relation and its inverse, plus at least two per datum to construct the master file, plus two for each computer word of data (numbers, alphanumeric strings, etc.) required to represent the contents of each datum, making a total for one binary link between two data, each containing two computer words, of twenty-four links. Using Cylinders, the requirement is two and one-half links for each datum, counting a tag field as one half of a link field, two and one half for each binary relationship represented, and no extra links for the contents of a datum because of the block structure. The total requirement for the case cited above is seven and one-half links, which is less than one third of the IPL-V requirement and the comparison improves as more words are needed for each datum.

15.2.4. Examples of Use

An example of a binary relational structure of a rather simple kind arises in the machine representation of line drawings involving points and straight lines as elements. If points are taken as the data, then lines can be represented as binary relations between the data points, and a

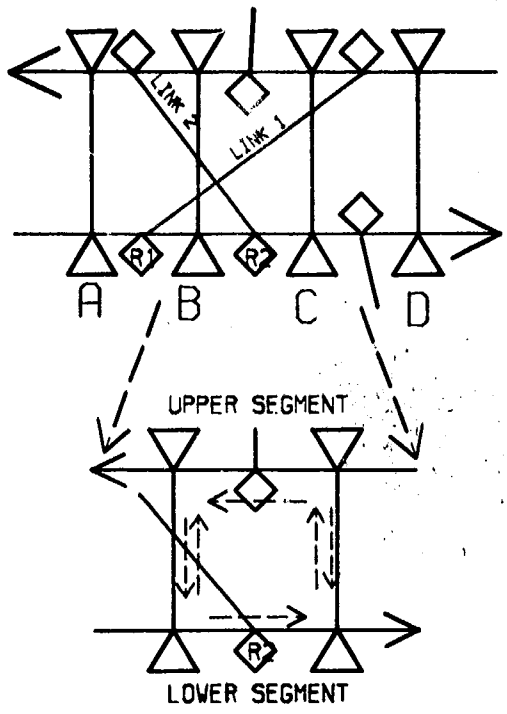


Fig. 15.13. The subring structure based on 'seams'. In the upper structure links 1 and 2 represent R1(a,c) and R2(b,a) respectively according to the binary relation scheme described in the text.

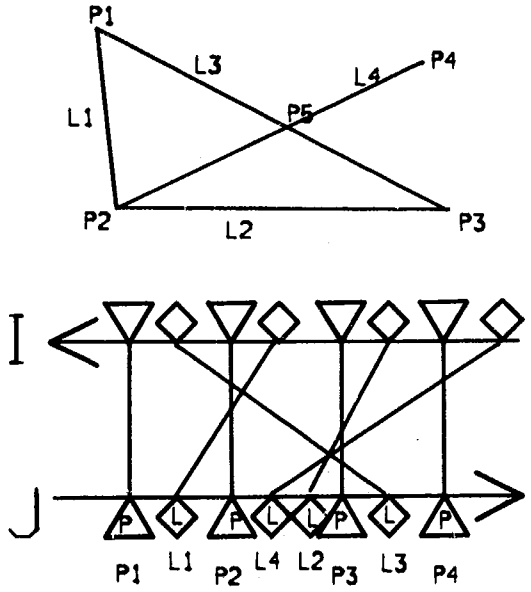


Fig. 15.14. A typical structure.

structure of the sort depicted in Fig. 15.14 results. This structure represents only the connections of the points and can stand for any pattern consisting of a triangle with a fourth line ending on one of its vertices. Supplying coordinate values for the points realizes a particular example.

As an example of the flexibility of cylinders, Fig. 15.15 shows a completely compatible extension of the scheme in Fig. 15.14 which allows the simultaneous representation of line-to-line relations, such as intersection, and of point-to-point relations including the line definitions in terms of point pairs. At the K and L levels, a second cylinder is formed, having as seams the lines which appear as relational links at the upper level. Any points which are defined by line intersections, such as point 5 in the diagram of Fig. 15.14, can appear as relational links at the lower level, but as seams at the upper level, thus becoming capable of serving to define further lines and so on.

Because points which arise from line intersections depend for their coordinates upon the other points which define the intersecting lines, this dependence is important to any program which operates upon such data structures and must be somehow marked, without destroying the identifiability of the datum as a point. This is easily done, e.g., by using the upper subfield of the tag in CYL6, marking it distinctly from the corresponding subfield for other points but keeping the lower portion the same. In this way the tags themselves can be made to represent a type of classification upon the data with no further burden upon the structure or storage requirement.

15.2.5. Summary

A data structure concept which is an adaptation of ring structures has been developed and described. It is highly compact, compared with conventional lists, possessing an advantage of typically a factor of three in storage requirements, and it is highly flexible in use, since its definition includes most other ring systems, but the elimination of almost all unnecessary system conventions, and a novel employment of the ring structure, tend to give Cylinders an advantage in compactness and flexibility in this field as well. It is now being actively employed in programming for artificial intelligence and its further potential thereby being explored.

P. Weston

15.3. Investigation of Fundamentals of Nonlinguistic Cognition

In this section, we describe an attempt to distill the mass of requirements for a cognitive system down to certain essential problems. Of fir t importance is the tentative conclusion that the representation of meaning of a communication must be based upon knowledge of the natural constraints between objects and relations, that is, upon a world model or the "eidology", according to the terminology introduced in "Cognitive Memory: An Epistemological Approach to Information Storage and Retrieval" (project proposal).

This is the direction which Chomsky takes in "Aspects of the Theory of Syntax" by supplying property markers. We feel, however, that Chomskian markers are far too inflexible as a mechanism for representing a world model. Rather, it appears that an approach guided by the philosophy

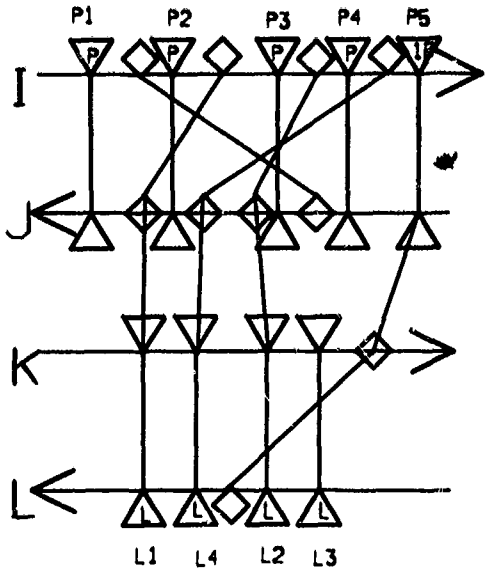


Fig. 15.15. An extension of the scheme of Fig. 15.14 that allows intersection points to be represented as binary relations between lines in a cylinder at one level while the same lines are simultaneously represented as relations between points in a cylinder at another level. P5 is given a distinct tag value, lp, because it must be distinguished as a dependent quantity, and is not a freely variable point.

(OVERLEAF BLANK)

of cognition expounded by Lenneberg in "Biological Foundations of Language" is much better suited to our task.

The principal message of Lenneberg, as it applies to this investigation, is that cognition is based upon a nonlinguistic discrimination process with verbalization as a largely extraneous (or, certainly, secondary) appendage.

In the same frame of concepts, a widely-accepted model for the description of cognitive processes contains two main transformational components, that is: 1) a mapping from the environment through percepts to the eidology; 2) a mapping from the eidology to the linguistic description, as well as its inverse. It is apparent that the two mentioned mappings "environment → eidology" and "eidology → language" have distinct characteristics which warrant a separate analysis. Furthermore we feel that the insight gained through the analysis of the perceptual mapping may greatly facilitate, in due time, the study of the linguistic mapping.

We were accordingly led to attempt to synthesize and study a very simplified model of a concept-discrimination process. The percepts chosen as stimuli (input) are stylized versions of "scenes" of an anthropomorphic environment or "real world" containing representations of shapes, colors, sizes, positions, etc. (i.e., visual scenes transformed so that these properties need not be graphically discriminated). The "real world" assumed as the underlying matrix of the percepts bears no character of necessity, but is only chosen for its convenience in view of an empirical test of the results of the investigation.

The final objective of this study is the development of a procedure for cognition, namely an analysis mechanism which converts a given scene into a description thereof, this description being partially classificatory and partially relational. It must be noted, however, that the analysis problem cannot be tackled prior to the precise definition of the "universe" upon which the cognitive mechanism is to operate. This clearly leads to the synthesis problem, namely to the description of the model which distinguishes admissible from inadmissible scenes. We can therefore distinguish a "generator model" and a "cognitor model" (usually contained within the former), which, roughly, play the roles of speaker and hearer, respectively, according to the well-known linguistic terminology.

The crux of the problem is therefore reduced to the specification of the nature of the "generator model" and of the embedded generative device. Intuition would suggest that the grammatical constraints should be rather weak and that the interdependence of the "details" should be rather strong, which is equivalent to strengthening the semantic constraints. How to express the weakening of the grammatical rules, or even more, the strengthening of the semantic texture is presently unclear. It appears practical and intuitively appealing, however, to express the constraints on the choice of details (i.e., the constraints on the rewriting rules) as determined by some broad properties of the remainder of the scene, instead of by the specific details occurring elsewhere in the scene itself.

The tentatively outlined structure is reflected, although in a somewhat incomplete form (which leaves the door open to some learning

capability), in the cognitor model. Upon presentation of scenes (percepts), it is necessary to disambiguate the communication by successively extracting properties, and attempting to infer the description of finer details on the basis of accumulated inexact data from the "world-model." The disambiguation will proceed from the general toward specific although backtracking (changing a guess) is a necessary provision, under the guide of a currently computed likelihood function (the strategy being very much analogous to the one used in sequential decoding). For example, the time and general location of a scene will first be guessed; these choices help to restrict the probable kinds of objects and their positions, etc.

At the present stage of the work, we wish to answer:

1) Which is an adequate structure for the implementation of the generator model (and, consequently, of the cognitor model)?

2) How to proceed to construct a description of a particular event (communication) as it is disambiguated?

F. P. Preparata
S. R. Ray

15.4. Cognition and Heuristics

The formation and use of concepts is assumed to be the main feature of cognitive systems. Research has been made towards mechanisms capable of acquiring concepts. As a working hypothesis, the following definition by Church (1956) has been adopted: A concept is a decision rule to be used to decide whether or not a given object is a member of a certain set of objects to which the name of the concept applies.

It is assumed that the acquisition of concepts must be motivated and controlled by the criterion of usefulness.

As a method of study, it has been decided to work on the outline of a tentative computer program for the solution of a specific class of problems within a simple universe. The class of problems chosen is the one of assigning 2-dimensional objects to locations on a plane with the goal of minimum connection cost, if these objects have to be connected in a given manner, and the connections cost is proportional to the distance to be bridged. The layout of etched wiring boards or the placement of machines in a factory are practical examples of this problem.

There does not exist a solution algorithm for these problems, except, in some cases, complete enumeration of the solution space, which is unfeasible for more than, say, 10 objects.

Restricted permutations are the most effective methods today. Unfortunately, they are qualitatively and quantitatively restricted to only the simplest assignment problems.

The point in these problems is to achieve a match between the topological structure of the connected objects and the geometrical pattern of the available spaces. Clearly, this involves the ability to recognize topological structures as well as geometrical shapes.

Work is now in progress on the definition of principles and procedures which would enable a system to learn and apply simple topological and geometrical concepts, as "chain, tree, ring," etc., and "corner, square, channel," etc.

As a sideline, a 1000-statement FORTRAN program named COBSOP (Connectivity Based Sequence of Placement) has been written. It serves for the study in more detail the handling of topological data and the degree of recognition obtainable with fixed, built-in routines. In addition, the program could be used as a tool for solving actual assignment problems.

H. Bielowski

15.5. Machine Architecture for Information Retrieval

15.5.1. Introduction

The object of this report is to describe an information-retrieval system which operates in a highly parallel fashion in an associative mode. By associative it is meant that a particular item is located in the data base by content addressing. In this process, an entire record of information is located using a known part of the record as the search criteria. As an elementary example, the record "John's car is bright blue" could be located or retrieved by specifying the words "blue" and "car." This method of data interrogation is desirable in all information-retrieval problems. The notion of addressing by contents enables us to avoid the artificialities of location assignment and frees us to a large extent from such local considerations as scanning, searching and counting; the operation is parallel due to the fact that the entire content of the memory is searched simultaneously. This is accomplished by distributing a sufficient amount of logic throughout the memory. This allows each storage location or "cell" to act, to a certain degree, as a small, independent, processing unit.

The associative memory processor described below is a logical

outgrowth of the organization proposed by Lee and Paull. [14] This previous organization was thoroughly investigated and it was found somewhat deficient in two areas. First of all, the general operation of the processor was found to be inefficient in terms of the number of steps (therefore the amount of time) and the amount of programming required to perform a data search. Secondly, but most important, it was found that certain types of searches which are extremely important to information retrieval could not be perfomed at all. The next section briefly describes the organization and operation of the processor presented by Lee and Paull. The shortcomings of this system and qualities which are desirable in a processor of this type are also presented. The final section describes in brief an associative memory processor which alleviates the problems presented and therefore facilitates a more efficient information retrieval system. Examples are given throughout to clearly explain the operation.

15.5.2. Investigation of the Processor of Lee and Paull

The machine proposed by Lee and Paull consists of an associative memory connected to a conventional, stored-program computer. The primary function of the computer is to act as a control device for the associative memory. It also must be able to store and execute the search programs and temporarily store data which is retrieved from the memory. The data base which is to be interrogated is stored in the associative memory.

The basic memory is a linear array of small, identical, sequential-state machines called cells. Each cell contains both a number of binary storage elements and a sufficient amount of logic to enable it

to perform the functions of the cell. The storage elements of the cells are of two types: the symbol elements and the activity elements. The symbol elements are those in which the information or "symbol" of the cell are stored. The activity elements are used as bookkeeping tags to keep track of particular cells during the operation of the machine. Each cell in the array is connected to a set of common bus lines, which include the input leads, the output leads and the various control leads which provide the cells with commands. A cell C_i also has the ability to transfer its activity status to either of its immediate neighbors, C_{i+1} or C_{i-1}. The contents of every cell (the symbol and/or the activity) can be matched against the pattern presented by the control computer on the input bus lines. In this way all cells containing a particular bit pattern can be located and tagged.

Information is stored in the computer as strings of symbols with each symbol stored in a different cell. (Strings can be of any length as long as there is room to store them.) Consecutive symbols of a string are stored in consecutive cells. Therefore if a string contains n symbols and the first symbol of the string is stored in cell C_i, then the second symbol is stored in C_{i+1}, the third in C_{i+2}, and so on with the nth symbol being stored in the C_{i+n} cell. The location of a string is completely arbitrary since the cells do not have addresses, but the order of the cells in which consecutive symbols are stored is very important. A special symbol α_o is provided for the beginning of a string. Parts of strings or parameters are distinguished by storing another special symbol, β, after them. Therefore if a string consisted of the parameters XY, OW, PHD, it would be

stored in the cell memory in the following manner: α_oXYβOWβPHDβ, with α_o stored in cell C_i, X in C_{i+1}, Y in C_{i+2}, and so on.

Identification and retrieval of a string or a particular set of strings are accomplished by using a particular parameter as the search criteria. For instance, if it were desired to locate all strings which contained the parameter OW the search would proceed in the following way. First, every cell in the memory would match its contents against the input bus lines which would contain the pattern β. This denotes the beginning of the parameter βOWβ. If a cell has a successful match it will set an activity bit. At this point all cells containing β have an activity bit set. Next, all cells with activity bits set are told to propagate this activity to their neighbor with a higher index (to the right) and then reset their own activity. Now all cells compare their contents with the input pattern consisting of a symbol equal to 0 and a set activity bit. Only those cells which contain a set activity and the symbol 0 are allowed to keep their activity set. All other previously-active cells are reset. At this point only those cells which are at the beginning of parameters and contain the symbol 0 are active. Again the activities are propagated to the right, and search is made for all active cells containing the symbol W, with all other activities being reset. Now only those cells containing W which is the second symbol of all parameters having 0 as the first symbol are active. Finally, the activity is again propagated to the right and a search is made for all active cells containing the symbol β. At this point only those strings which contain the parameter OW have an active cell. This cell is the cell

containing β which follows OW.

This searching process is a very efficient one. It can be thought of as eliminating useless information rather than a searching process for useful information. Although the search is done serially by character, it is a parallel operation conducted in each of the strings simultaneously.

Although at this stage all the strings containing the parameter OW have been isolated, more processing is required before these strings can be retrieved. First, a priority system must determine which of the strings of the set is to be retrieved first. After this string is found, the activity present in this string must be transferred to the cell containing the α_0, head cell, of that string. This is accomplished by propagating the activity to the neighboring cell with lower index (the cell to the left) and matching for a cell which is active and contains the symbol α_0. If the match is successful, the head cell of that string has been found. If it is unsuccessful, the process is repeated until the head cell is found. From this point, the characters in the string are read out sequentially, one at a time, by successively propagating the activity to the right neighbor and ordering the active cell to place its contents on the output lines. This process, of course, is non-destructive.

This example introduces two of the shortcomings of this processor. First, the priority system of this machine only allows one string of the set to retain an active cell. This string is the one to be retrieved. Since they no longer contain an active cell the remaining strings in the set must again be identified by another search in order to be again eligible for

retrieval. Provision is made to eliminate previously retrieved strings from subsequent searches, but still the same search operation must be performed as many times as there are strings in the set to be retrieved. These repetitive searches require additional operating time during which no really useful processing is being accomplished. It would be desirable to be able to retain the active status of all the strings of a set during the retrieval of the entire set. This would require that only one search, the initial search, be made. The second deficiency to be noted here is the waste of processing time needed to find the head cell in a string chosen for retrieval. The activity must be propagated to the left and a match be performed as many times as the number of cells between the head cell of the string and the active cell which resulted from the search. If this number is large, the time involved to complete this process becomes very costly, because the rest of the memory must stand idle until its completion. This suggests that it would be advantageous if all the cells in a string could communicate directly with the head cell without the necessity of involving other cells.

Additional areas in which this processor and all other similar associative memory processors [15, 16, 17] designed for information retrieval are deficient can be placed in two general categories. The first are the operations the machine can execute but does not do so efficiently. The two mentioned above fall in this class.

Another problem would be that of identifying the set of all strings which contain two or more particular parameters. If the parameters are stored in a particular order, and they are stored in contiguous groups of cells, the

searching process is the same as above, e.g., A and C are the parameters which the strings must contain, and they are stored in all strings of the set containing A and C in the form βAβCβ. If one or both of these requirements is not met (as is usually the case) then the searching process becomes quite complicated and requires a great amount of processing time.

Still another problem which has not been handled efficiently is that of repacking the strings in the memory after information has been eliminated. This is necessary so that all the empty cells in the memory can be utilized to store new strings of arbitrary lengths. The methods of gap closing proposed by the various authors involve programs which are extremely inefficient with regard to the amount of processor time it takes to repack the memory.

The second category includes those types of operations which are necessary for an efficient information retrieval system [18] but cannot be performed by the type of associative processor suggested by Lee and Paull and its descendents. [16,17]

The organization described above identifies the set of all strings which contain the parameters of the search criteria exactly. It cannot perform a threshold search. A threshold search can be described as the identification of all strings which contain at least a certain number of the parameters used for the search. For example, it might be desirable to locate all strings which contain any 3 or more of 5 particular parameters.

A second type of search which would be useful would be a weighted threshold search. In this identification procedure, the presence of a

particular parameter in a string may be of more or less value than that of other parameters. In this operation each parameter used in the search would be given a weight or value corresponding to its importance. If a string contains a particular parameter, this value is recorded and added to the results of previous and subsequent parameter searches. After all parameters have been searched, only those strings which contained a total search value greater than a certain amount would be of interest for retrieval. Both types of searches described above imply that the strings should be capable of storing the results of many searches. No provision is made for this in the associative memory processors mentioned above.

The following section is a brief description of an associative memory processor which will alleviate the problems outlined above and facilitate a more efficient information retrieval system.

15.5.3. A New Associative Memory Processor

The structure of this associative memory processor is essentially the same as the configuration proposed by Lee and Paull. The basic differences lie in the memory cell. This allows an entire string of cells to act together as an independent processing unit.

The basic configuration consists of general purpose computer acting as a control device for an associative memory. The computer stores and executes the search programs, issues commands to the control leads, and temporarily stores the data which is to be entered into or retrieved from the associative memory. The memory stores the data base which is to be interrogated.

As in the previous model, the memory consists of a linear array of small sequential state machines called cells. Each cell is connected to a common set of bus lines. These are the input leads, output leads and control leads. These lines allow the control computer to communicate with all the cells simultaneously (if the propagation time along the lines is ignored).

All cells in the memory are identically the same. Each cell contains both storage elements, binaries, and a certain amount of logic which enables it to perform the various functions required.

The storage elements are divided into four different fields. A six-bit field stores the character or symbol of the cell. This is the information register. There is a one-bit activity field. This bit is used as a bookkeeping tag to keep track of certain cells during processing. Another bit is used to indicate if a cell is empty or if its symbol contents is no longer considered important. The last bit is the α_o bit. The presence of this bit allows a string of cells a high degree of processing power.

A cell may function in one of two states. The state of a particular cell is determined by the condition of its α_o bit. If the α_o bit is set the cell acts as the head cell of a string. In this state the cell is able to receive count pulses which can be emitted from all the cells of higher index up to the next head cell. The binary representation of the number of these pulses are stored in the head cell's symbol register, which now acts as an accumulator. This concept of the function of a head cell is very important. It allows the results of successive searches to be

stored in each string and eliminates the necessity to propagate an activity to a cell of lower index (to the left). It also enables the memory to perform both a threshold search and a weighted threshold search with equal ease. These points will be explained clearly in the example at the end of this section.

Each cell in the memory can communicate directly with other cells in three different ways. First, a cell can transmit its set activity to its neighbor with a higher index (to the right). Secondly, as mentioned above, an active cell in a string can transmit a pulse directly to the head cell of its string. This pulse is sent along a special gated network called the routing line. This is a one directional path which passes through all the cells in the memory. Part of the routing line is shown in Fig. 15.16. One head cell and two symbol cells are displayed. Last of all, each of the cells has the ability to shift the contents of all its registers to its neighbor on the right. This is used when repacking the memory and can also be used to load or unload the entire contents of the memory automatically.

During a match, the contents of every cell is simultaneously compared with the pattern presented on the input lines. If a cell's contents matches this pattern, the activity bit of that cell is set. "Don't care" conditions can be specified in any of the bit positions so that a certain set of bits may be examined at any one time.

Information is stored in the memory as strings of characters. Each character of the string is stored in a distinct cell. The first cell in the string is always a head cell. The special symbol γ always precedes and

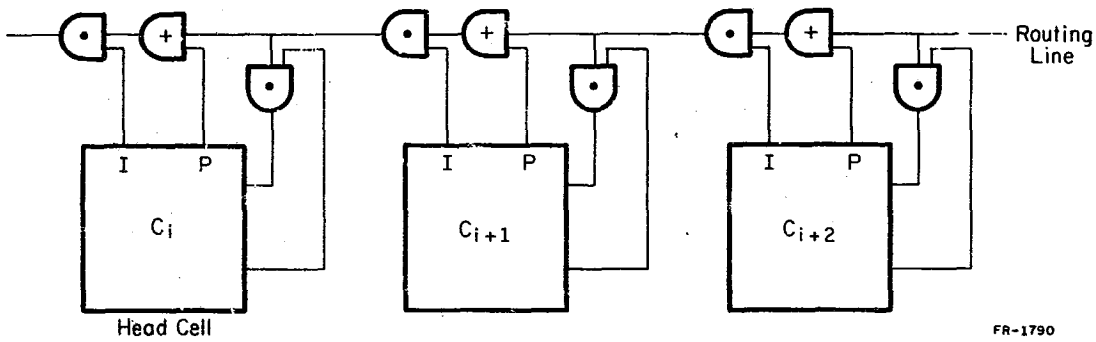


Fig. 15.16. A typical organization.

(OVERLEAF BLANK)

follows the parameters of a string and the last two symbols in a string must both be γ. As an example, the string containing the parameters XY, OW, PHD, PCC, and AF would be stored in the following way:

$$---\alpha_o\gamma XY\gamma OW\gamma PHD\gamma PCC\gamma AF\gamma\gamma---.$$

As an example of operation, consider the problem of locating all strings which contain the parameters XY and AF. The process for locating a particular parameter is the same as in the example in the previous section. First, XY is located everywhere in the memory. At this point, only those strings which contain XY have a cell with its activity set. A control bus now tells all cells with their activities set to place a pulse on the routing line. These pulses are now recorded in the head cells of all the strings containing the parameter XY. Each of these head cells now contain a binary count of one in their symbol registers. The activities are reset and a search is now made to find all the occurrences of the parameter AF in the memory. After these are found only those strings which contain AF have a cell with its activity set. Again, a control bus orders all active cells to place a pulse on the routing line. At this point, only the head cells of the strings which contain both XY and AF have a binary count of 2 in their symbol registers. All activities are again reset. Now the head cells of the set of strings which contain both parameters can be activated by matching the contents of every cell with the appropriate pattern on the input lines. At this point, a priority system can be used which resets all but one active head cell. The search for the parameters XY and AF will not have to be made again, because the results of the initial search are still

stored in the symbol registers of the head cells. These cells can be activated again by a simple match operation. Before a string is read out, the bits of its symbol register are reset so that it will not be considered again in the priority search.

It is evident from this example that many processing steps and therefore much time has been saved by allowing the head cells to both store the results of consecutive searches and communicate directly with the cells in their strings.

It is apparent that a threshold search would be carried out in exactly the same way. The threshold value would be used when matching for head cells which contain a particular stored number. A weighted threshold search would be performed by instructing the active cells resulting from a parameter search to issue a number of pulses equal to the relative importance of the parameter. After all searches are completed, the head cells would be tagged if they contained a value greater than a certain number.

A. Kisylia
R. T. Chien

15.6. Studies of the Mathematical Theory of Cognition

15.6.1. On the Forms of Equations Associated with Inductive Inference Computers

As indicated by H. Von Foerster in various publications such as "Memory and Inductive Inference" and "Memory Without Record" (The Anatomy of Memory) to build a computer that learns from experience, it has to be "pliable" or, more accurately, it must be an adaptive computer. Whatever other aspect

built into the computer it should be able to make "inductive inferences." This implies that the system should be able to compute future events from past experience. Because the next sentence as given in "Memory Without Record" is pertinent and crucial to what follows, I place it in quotes, "It is clear that only a system that has memory is capable of making inductive inferences because, from a single time-slice of present events, it is impossible to infer about succeeding events unless previous states of the environment are taken into consideration." If, as is often done, we think of the "brain" as a large number of neurons which are interconnected in some way to form a complicated network, the stimulus-response equations should embody position coordinates, a time coordinate, and what is crucial, terms with time retardation which would give the computer the essential characteristic previously mentioned, namely, the ability to predict future events on the basis of a knowledge of past states. That is, to give an equation or a system of equations with initial conditions--that is, the state at this moment, say $t = 0$, and no retarded arguments in time--is insufficient to characterize an inductive-inference computer.

It follows that systems that are capable of "predicting" or "learning from experience" ought to be characterized by equations of the form

$$\begin{aligned} \dot{x}(t) &= f(t,x(t),x(t-\tau)),\ t > \tau \\ x(t) &= g(t),\ 0 \leq t \leq \tau, \end{aligned} \tag{1}$$

where $\tau > 0$ and $g(t)$ characterize the "past history." More, generally, they should be governed by a system such as

$$\begin{aligned}
\dot{x}_1(t) &= f_1[t,x_1(t),x_1(t-\tau_{11}(t)),\ldots,x_1(t-\tau_{1m}(t)),\ldots,\\
&\qquad x_n(t),x_n(t-\tau_{n1}(t)),\ldots,x_n(t-\tau_{nm}(t))],\\
\dot{x}_2(t) &= f_2[t,x_1(t),x_1(t-\tau_1,(t)),\ldots,x_1(t-\tau_{1m}(t)),\ldots,\\
&\qquad x_n(t),x_n(t-\tau_{n1}(t)),\ldots,x_n(t-\tau_{nm}(t))],\\
&\qquad\qquad \ldots\\
\dot{x}_n(t) &= f_n[t,x_1(t),x_1(t-\tau_{11}(t)),\ldots,x_1(t-\tau_{1m}(t)),\ldots,\\
&\qquad x_n(t),x_n(t-\tau_{n1}(t)),\ldots,x_n(t-\tau_{nm}(t))],
\end{aligned} \tag{2}$$

or, more shortly,

$$\begin{aligned}
\dot{x}_i(t) &= f_i[t,x_1(t),x_1(t-\tau_{11}(t)),\ldots,x_1(t-\tau_{1m}(t)),\ldots,\\
&\qquad x_n(t),x_n(t-\tau_{n1}(t)),\ldots,x_n(t-\tau_{nm}(t))],\\
i &= 1,2,\ldots,n, \quad t_0 \le t \le B, \quad \tau_{ij}(t) \ge 0
\end{aligned} \tag{3}$$

and the initial conditions (constituting past history)

$$x_i = x_{i0}(t), \tag{4}$$

defined on the set E_{t0} consisting of the point t_0 and all those differences $t-\tau_{ij}(t)$ $(i=1,2,\ldots,n;\ j=1,2,\ldots,m)$ which are less than t_0 for $t_0 \le t \le B$.

The mathematics dealing with such systems has, to a certain extent, been developed. It is far from complete, however. Nevertheless, certain steps in the direction of an inductive-inference computer can be taken by

considering certain functions f_i and studying their solutions. The forms of the function f in (1) and f_i in (2) would play a crucial role. What their structure should be is not too apparent at the present.

An examination of the literature indicates that my surmise about the form of the Eqs.(1) that go with a possible prediction theory is substantiated in at least one case by an article by Grossberg.[19] In this paper, the author considers a system of the form

$$\dot{X}(t) = AX(t)+B(X_t)X(t-\tau)+C(t), \quad t \geq 0 \tag{5}$$

where $X = (x_1,\ldots,x_n)$ is non-negative, $B(X_t) = [B(t)]$ is a matrix of non-negative functions of X(w) evaluated at past times $w \in [-\tau, t]$, and $C(t) = (I_1,I_2,\ldots,I_n)$ is an input function.

Specifically, the system studied was

$$\begin{aligned} \dot{x}_i(t) &= -\alpha x_i(t)+\beta \sum_{k=1}^{n} x_k(t-\tau)y_{ki}(t)+I_i(t), \\ y_{jk}(t) &= p_{jk}z_{jk}(t)\Big(\sum_{m=1}^{n} p_{jm}z_{jm}(t)\Big)^{-1}, \\ z_{jk}(t) &= [-uz_{jk}(t)+\beta x_j(t-\tau)x_k(t)]\ \theta\ (p_{jk}), \end{aligned} \tag{6}$$

which can be written in vector form

$$\underset{\sim}{U}(t) = f(t,\underset{\sim}{U}(t),\underset{\sim}{U}(t-\tau)), \tag{7}$$

with

$$\underset{\sim}{U} = (x_1, x_2, \ldots, x_n, z_{11}, z_{12}, \ldots, z_{n,n-1}, z_{nn}),$$

$$F = (f_1, f_2, \ldots, f_n, f_{11}, f_{12}, \ldots, f_{n,n-1}, z_{nn}),$$

$$f_i = -\alpha x_i + \beta \sum_{k=1}^{n} x_k(t-\tau) p_{ki} z_{ki} \sum_{m=1}^{n} p_{km} z_{km}^{-1} + I_i,$$

and

$$f_{jk} = [-u z_{jk} + \beta x_j(t-\tau) x_k] \theta (p_{jk}).$$

Equation(7) corresponds to the system given in Equation(3). By restricting its function somewhat, a machine based on a system like (7) can be taught to predict an event B whenever the event A occurs. This is phrased in another way by saying that we wish to teach the machine the list AB, or, in terms of an idealized human subject H, we wish to teach H the list of letters AB.

It would appear that a system such as (3) is involved in the problem of building a teaching machine. What needs to be specified are the functions f_i which would incorporate such properties as remembering, correcting errors, improving with practice, et.

E. J. Scott

15.6.2. On a Class of Nonlinear Property Filters

15.6.2.1. Introduction.

In a previous report [20] and article [21] a set of elements q of a set L, a stimulus function $\sigma(q)$, a response function $\rho(q)$, and an interaction function $K_1(p,q)$ are considered to be interrelated in such a manner that the resulting equation is a linear integral equation. Both

interaction and action in linear systems, as well as discrete linear systems and those with uncountable number of elements were considered and applied to certain specific situations.

It is the purpose of this report to modify the interaction function so that the resulting stimulus-response equation is a nonlinear integral equation. The result should be more in keeping with the fact that in nature we are in general dealing with nonlinear phenomena and such an equation or system thereof would be more representative than linear systems. As is well known, replacing linear systems, about which we know a good deal, with nonlinear systems, about which we know relatively little, presents great mathematical difficulties. We shall consider a class of nonlinear property filters leading to integral equations for which some mathematical theory has been worked out.

15.6.2.2. Interaction in Nonlinear Systems.

In order to maintain continuity we shall use most of the notation in the report. [20] Let $p = (x_1, x_2, \ldots, x_n)$ and $q = (\xi_1, \xi_2, \ldots, \xi_n)$ be any two points of an n-dimensional manifold $\mathcal{E}_n$ and $d\mu_q$ a volume element about the point q. Suppose that an element q of $\mathcal{E}_n$ is subjected to a stimulus $\sigma(q)$ and that the connectivity between each element q of $\mathcal{E}_n$ and all other elements in $\mathcal{E}_n$ has been defined. We denote by $\rho(p)$ the response to a stimulus at p. We shall now assume that the amount of stimulation received by p from q, instead of involving the response ρ linearly, takes it into account in a nonlinear manner. To be specific, we shall assume that the amount of stimulation received by p from q is

$$K_1[p,q;\rho(q)], \tag{1}$$

where K_1 is the so-called interaction function. It should be mentioned that one of the problems here is the delineation of the form of K_1 if it is to reflect the response of an actual physical system. Some experiments here, if possible or feasible, would be of value in determining the analytic character of K_1. Under these assumptions, the total amount of stimulation at p as q encompasses $\mathcal{E}_n$ will be given by the nonlinear integral equation

$$\rho(p) = k\sigma(p)+\lambda\int_{\mathcal{E}_n} K_1[p,q;\rho(q)]d\mu_q,\ p,q \in \mathcal{E}_n, \tag{2}$$

where k is a constant amplifying factor for $\sigma(p)$ and λ is an amplifying factor reflecting the stimulation at p due to the totality of elements q belonging to $\mathcal{E}_n$. No general theory for the solution of Eq.(2) with an arbitrary nonlinear kernel K_1 exists. Instead, certain classes of nonlinear integral equations, especially those tied to physical problems, have been examined. For some of these existence and uniqueness theorems have been developed as well as methods of determining solutions. One such class is that discussed in the paper by Hammerstein. [22]

15.6.2.3. A Class of Nonlinear Filters.

Let us consider the class of nonlinear integral equations

$$P(p)+\int_{\mathcal{E}_n} K(p,q)F[q,P(q)]d\mu_q = 0,\ p,q \in \mathcal{E}_n \tag{3}$$

of the Hammerstein type. We assume that the kernel $K(p,q)$ belongs to the class L_2, which implies that the functions

$$\left(\int_{\mathcal{E}_n} K^2(p,q)\,d\mu_q\right)^{\frac{1}{2}} = \alpha(p), \tag{4}$$

$$\left(\int_{\mathcal{E}_n} K^2(p,q)\,d\mu_p\right)^{\frac{1}{2}} = \beta(q), \tag{5}$$

exist almost everywhere in $\mathcal{E}_n$, and their squares are Lebesgue integrable, i.e.,

$$\|K^2(p,q)\| = \int_{\mathcal{E}_n} \alpha^2(p)\,d\mu_p = \int_{\mathcal{E}_n} \beta^2(q)\,d\mu_q \leq M^2,$$

where $M > 0$.

Equation(2) for a kernel of the type we are considering would have the form

$$\rho(p) = k\sigma(p) - \int_{\mathcal{E}_n} K(p,q)\mathcal{F}[q,\rho(q)]\,d\mu_q = 0, \tag{6}$$

where

$$\lambda K_1[p,q;\rho(q)] = -K(p,q)\mathcal{F}[q,\rho(q)]. \tag{7}$$

A substitution

$$P(p) = \rho(p) - k\sigma(p) \tag{8}$$

transforms (6) into

$$P(p) + \int_{\mathcal{E}_n} K(p,q)\mathcal{F}[q,P(p)+k\sigma(p)\,d\mu_q = 0, \tag{9}$$

which is of the form of Eq.(3).

A principal analytical method of solving (3) or (9) is to employ the classical method of successive approximation scheme by setting

$$P_0(q) \equiv 0 \tag{10}$$

and determining successive functions by means of the relations

$$P_{n+1}(p) = -\int_{\mathcal{E}_n} K(p,q)F[q,P_n(q)]d\mu_q, \quad p,q \in \mathcal{E}_n, \tag{11}$$

$n = 0,1,2,3,\ldots$. The convergence of successive functions $P_i(p)$ is not assured, of course, unless certain conditions are satisfied. A generalization of the proof in Tricomi [23] for the one-dimensional case to an n-dimensional space shows that the sequence

$$P_0(p),P_1(p),P_2(p),\ldots,P_n(p),\ldots$$

converges almost everywhere to a solution of (11) if:

(a) $\int_{\mathcal{E}_n} K^2(p,q)d\mu_q = \alpha^2(p)$, exists everywhere in $\mathcal{E}_n$,

(b) $F(q,v)$ satisfies the Lipschitz condition of the form

$$|F(q,v_1)-F(q,v_2)| < \gamma(q)|v_1-v_2|,$$

(c) $F(q,0)$ belongs to L_2,

and

(d) $\int_{\mathcal{E}_n} \alpha^2(p)\gamma^2(p)d\mu_p = M^2 < 1$.

These conditions are fairly stringent. Nevertheless, they are met by a large number of functions. Since these are sufficient conditions and not necessary, the method can still be employed, but the sequence of functions must be examined for convergence and satisfaction of the equation. In

cases where the evaluations are tedious and involved, numerical procedures may be employed.

As an example, consider the equation

$$\rho(x) = \int_0^1 K(x,\xi)(1+\rho(\xi))^2 d\xi,$$

where

$$K(x,\xi) = x, \xi < x$$

$$= \xi, \xi \geq x.$$

We define the iteration process by the sequence

$$\rho_0(x) \equiv 0$$

$$\rho_{n+1}(x) = \int_0^1 K(x,\xi)[1+\rho_n(\xi)]^2 d\xi, \quad n=0,1,2,\ldots .$$

For n=0 we have

$$\rho_1(x) = \int_0^1 K(x,\xi)d\xi = x\int_0^x d\xi + \int_x^1 \xi d\xi = x\xi\Big|_0^x + \frac{\xi^2}{2}\Big|_x^1$$

$$= x^2 + \tfrac{1}{2} - \tfrac{1}{2}x^2 = \tfrac{1}{2}(x^2+1).$$

For n=1,

$$\rho_2(x) = \int_0^1 K(x,\xi)[1+\tfrac{1}{2}(\xi^2+1)]^2 d\xi = (1/2)^2 \int_0^1 K(x,\xi)(\xi^2+3)^2 d\xi$$

$$= (1/2)^2\left\{x\int_0^x (\xi^2+3)^2 d\xi + \int_x^1 \xi(\xi^2+3)^2 d\xi\right\}$$

$$= (1/2)^2\left\{(1/30)x^6 + (7/2)x^4 + (9/2)x^2 + (31/6)\right\}.$$

The iterates $\rho_3(x)$, $\rho_4(x)$, etc., are obtained similarly.

The crucial element in making use of this theory, as has been mentioned, is the determination of the nature of the kernel $K(p,q)$.

E. J. Scott

15.7. References

1. D. G. Hays, "Automatic Language-Data Processing," in H. Borko, ed., Computer Applications in the Behavioral Sciences, Englewood Cliffs, N.J., Prentice-Hall, 1962.

2. D. G. Hays and T. W. Ziehe, "Studies in Machine Translation--10: Russian Sentence Structure Determination," RM-2538, The Rand Corp., Santa Monica, Calif.,(April 1960).

3. K. McConlogue and R. F. Simmons, "Analyzing English Syntax with a Pattern Learning Parser," Comm. of ACM 11, 687 (Nov. 1965).

4. H. Gaifman, "Dependency Systems and Phrase-Structure Systems," Information and Control 8, 304 (June 1965).

5. P. Weston, "Data Structures for Computations within Networks of Relations," in BCL Report 67.2, Biological Computer Laboratory, University of Illinois, Urbana, 126 pp. (1967).

6. A. Newell et al., Information Processing Language - V, Prentice-Hall, Englewood Cliffs, N.J., (1964).

7. McCarthy, et al., LISP 1.5 Programmer's Manual, MIT Press, Cambridge, Mass. (1962).

8. A. J. Perlis and C. Thornton, "Symbol Manipulation by Threaded Lists," CACM 3, 195 (1960).

9. J. Weizenbaum, "Symmetric List Processor," CACM 6, 524 (1963).

10. L. G. Roberts, "Graphical Communication and Control Languages," in Second Congress on Information System Sciences, Hot Springs, Virginia (1964).

11. K. C. Knowlton, "A Programmer's Description of L6," CACM 9, 616 (1966).

12. P. Weston, "Data Structures for Computations within Networks of Relations," in BCL Report 67.2, Biological Computer Laboratory, University of Illinois, Urbana, 126 pp. (1967).

13. J. Bouknight, Preliminary User's Manual CSL6 CSL7, Coordinated Science Laboratory, University of Illinois, Urbana (1967).

14. C. Y. Lee and M. C. Paull, "A Content Addressable Distributed Logic Memory with Applications to Information Retrieval," Proceedings of the I.E.E.E. 51, 925 (June 1963).

15. C. Y. Lee, "Intercommunicating Cells--Basis for a Distributed Logic Computer," Proc. Fall Joint Computer Conf. 22, 130 (December 1962).

16. R. S. Gaines and C. Y. Lee, "An Improved Cell Memory," I.E.E.E. Trans. on Electronic Computers C-17, 10 (January 1968).

17. J. N. Sturman, "An Iteratively Structured General Purpose General Purpose Digital Computer," I.E.E.E. Trans. on Electronic Computers EC-14, 2 (January 1968).

18. G. Salton, "Progress in Automatic Information Retrieval," I.E.E.E. Spectrum, 90 (August 1965).

19. S. Grossberg, "A Prediction Theory of Some Non-linear Functional-Differential Equations. 1. Learning of Lists," Journal of Mathematical Analysis and Applications 21, 643 (March 1968).

20. A. Inselberg and H. Von Foerster, "Linear Property Filters," Technical Report No. 2, Electrical Engineering Research Laboratory, Engineering Experiment Station, University of Ill., Urbana, Ill., 1962.

21. H. Von Foerster, Computation in Neural Nets," Currents in Modern Biology 1 (March 1967).

22. A. Hammerstein, "Nichtlineare Integralgleichungen nebst Anwendung," Acta Math. 54, 117 (1930).

23. F. G. Tricomi, "Integral Equations," Interscience Publishers, Inc., New York (1957).

(OVERLEAF BLANK)

16. SEISMIC DETECTION[†]

H. W. Knoebel B. D. Kirkwood

16.1. Seismic Vehicle Identification

Since the end of the previous Progress Report period, additional seismic and acoustic data has been recorded for use in a study of the relative contribution to the spectrum of roadway characteristics, speed-independent vehicle characteristics, and speed-dependent vehicle characteristics.

About 100 sets of correlated data as well as recordings of background noise were obtained using two different vehicles and two different road types. A 1956 Ford light truck and a 1966 Dart sedan were used over an unsurfaced dirt road in an isolated, partially-wooded area and over an old concrete pavement now bypassed by a new highway about one-fourth mile away.

Test runs were made in first, second, and third gears at 1000, 1500, 2000, and, where practical, 2500 engine revolutions per minute. Considerable effort was made to maintain constant engine speed during these runs for the purpose of implementing a partially-successful attempt to achieve high-resolution, speed-dependent spectra using only the fast-Fourier-transform data-processing technique.

To date, about 15 of the runs have been analyzed with the fast-Fourier-transform computer program using 500 12-bit samples per second. It is already evident that, in spite of the effort, engine speed varied sufficiently during the 16-second sample intervals to seriously broaden the speed-dependent components of the spectrum. This points up the need

[†]This work was supported by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB-07-67-C-0199.

to eliminate broadening effects of a nonstationary spectrum as would be done in the technique proposed by CSL for frequency tracking and with a variable sampling rate that adapts to the engine speed.

In addition to engine rotation rate, present spectra show evidence of propeller-shaft rotation, wheel-rotation harmonics, and intermodulation products. Frequencies above 60 Hz, where most of the firing rates (exhaust note) occur, have not been critically examined, because the two vehicles used are fairly-well muffled and do not couple much speed-dependent, high-frequency energy to the ground. In the few runs examined, the concrete pavement appears to attenuate the seismic high frequencies more than does the dirt road.