①

UNITED STATES AIR FORCE
WEAPONS LABORATORY
RESEARCH SCHOLAR PROGRAM

1983 -1984

MANAGEMENT AND TECHNICAL REPORT

DTIC
SELECTE
NOV 1 5 1985
E

CONDUCTED BY THE

SOUTHEASTERN CENTER FOR

ELECTRICAL ENGINEERING EDUCATION

(SCEEE)

WARREN D. PEELE

EARL L. STEELE

PROGRAM DIRECTORS, SCEEE

DR. JOHN UNGVARSKY

AFWL

SCEEE PRESS

85 11 14 075

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR. 85- 0906

6a NAME OF PERFORMING ORGANIZATION Southeastern Center for Electrical Engineering Education	6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research/XOT
6c ADDRESS (City, State, and ZIP Code) 11th & Massachusetts Ave. St. Cloud, Florida 32769		7b. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332

8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) XOT	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-82-C-0035

8c. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2301	TASK NO. D5	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

USAF WEAPONS LABORATORY RESEARCH SCHOLAR PROGRAM

12. PERSONAL AUTHOR(S)
Warren D. Peele, Earl L. Steele

13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) October 1984	15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

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

(See Attached)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL AMOS OTIS, Major, USAF, Program Manager	22b. TELEPHONE (Include Area Code) (202)767-4971	22c. OFFICE SYMBOL XOT

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.
All other editions are obsolete.

UNITED STATES AIR FORCE
WEAPONS LABORATORY
RESEARCH SCHOLAR PROGRAM

1983 - 1984

MANAGEMENT AND TECHNICAL REPORT

CONDUCTED BY THE

SOUTHEASTERN CENTER FOR
ELECTRICAL ENGINEERING EDUCATION
(SCEEE)

WARREN D. PEELE

DR. JOHN UNGVARSKY

EARL L. STEELE

AFWL

PROGRAM DIRECTORS, SCEEE

DTIC
ELECTE
1 5 1985

SCEEE PRESS

E

1983-1984 USAF/SCEEE
AIR FORCE WEAPONS LABORATORY RESEARCH SCHOLAR PROGRAM

conducted by

Southeastern Center for

Electrical Engineering Education

under

USAF Contract Number F49620-82-C-0035

MANAGEMENT AND TECHNICAL REPORT

submitted to

Air Force Office of Scientific Research

Bolling Air Force Base

Washington D.C.

by

Southeastern Center for
Electrical Engineering Education
(SCEEE)

October 1984

Accession For

NTIS GRA&I

DTIC TAB

Unannounced

Justification

By

Distribution /

Availability Codes

Avail and/or

Dist Special

A-1

1

UNITED STATES AIR FORCE

AIR FORCE WEAPONS LABORATORY RESEARCH SCHOLAR PROGRAM

1983-1984

PROGRAM MANAGEMENT AND TECHNICAL REPORT

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION
(SCEEE)

2

TABLE OF CONTENTS

1983-1984 AIR FORCE WEAPONS LABORATORY RESEARCH SCHOLAR PROGRAM

I. SUMMARY

The AFWL Research Scholar Program was initiated as a pilot program to provide new research scholars with one year appointments at the Air Force Weapons Laboratory.

Extensive mailings were made to technical departments at universities around the United States where programs of prime interest to the Weapons Laboratory were established. These included nuclear physics, radiation effects, electromagnetics, laser optics and related applied sciences.

Four scholars were appointed beginning in September 1983 and extending through September 1984, for 12 months duration.

Several technical papers were presented by the scholars during the year. The final technical reports on the scholar's work are included in this report.

This pilot program was judged a success by both the scholars and their laboratory associates. Their comments were solicited by questionnaire and are included. The scholars were judged to be beneficial to the laboratory. The opportunity of having new research people on a short term basis was felt to be very stimulating and worth while. Their interaction with the laboratory was very positive.

Overall, the scholars felt their experiences at the laboratory were constructive steps in their professional development.

II. INTRODUCTION

The Air Force Office of Scientific Research, the Air Force Weapons Laboratory, and the Southeastern Center for Electrical Engineering Education initiated a pilot U.S. Air Force Weapons Laboratory (AFWL) Research Scholar Program beginning in the fall of 1983. This new program was an adjunct effort to the U.S. Air Force Summer Faculty Research Program (SFRP) established by contract modification dated March 1983 under a special studies clause. This pilot program provided research opportunities for selected engineers and scientists holding a doctoral degree to work in residence at the Air Force Weapons Laboratory for a one year research period.

To be elgible, all candidates must be a U.S. citizen and have a Ph.D. or equivalent in an appropriate technical field. The scholars will be selected primarily from such basic and applied science and engineering fields as physics, particularly nuclear and laser physics, civil, electrical, aeronautical, nuclear, and mechanical engineering and also from applied mathematics and computer science.

The AFWL Scholar in this program had the following specific obligations:

1) To participate in advanced research programs at the Air Force Weapons Laboratory;

2) To prepare a report at the end of the one year appointment describing their research accomplishments. This report will be approved by the Air Force Weapons Laboratory;

3) To complete an evaluation questionnaire on the Air Force Weapons Laboratory Research Scholar Program.

III. APPLICATION INFORMATION

Qualified technical people who were interested in an appointment under this program were asked to file a formal application and supporting materials with the program director at SCEEE. Formal application forms and program announcements were widely distributed and also available from the SCEEE programs office. In this program SCEEE supports equal opportunity/affirmative action so that all qualified applicants received consideration without regard to race, color, religion, sex, or national origin.

The application deadline was July 15, 1983.

AIR FORCE WEAPONS LABORATORY SCHOLAR PROGRAM

AFWL SCHOLAR PROGRAM OBJECTIVES:

(1) To provide a productive means for scientists and engineers holding Ph.D. degrees to participate in research at the Air Force Weapons Laboratory;

(2) To stimulate continuing professional association among the research scholars and their professional peers in the Air Force;

(3) To further the research objectives of the United States Air Force;

(4) To enhance the research productivity and capabilities of scientists and engineers especially as these relate to Air Force technical interests.

PREREQUISITES FOR APPOINTMENTS: To be qualified for consideration as an AFWL Scholar in the fall 1983 program, the applicant must:

(1) Be a U.S. citizen;

(2) Be the holder of a Ph.D. degree, or equivalent, in an appropriate technical specialty;

(3) Be willing to pursue research work of limited time duration at the Air Force Weapons Laboratory.

Although it is anticipated that the research itself may be unclassified, the scholar must hold or be eligible for a Department of Defense SECRET clearance in order to insure access to work areas.

RESEARCH PERIOD: The period of this appointment is for one year at the Air Force Weapons Laboratory research site, Kirtland AFB, New Mexico beginning in the fall 1983 and ending during 1984.

FINANCIAL TERMS: The planned stipend for the AFWL Research Scholar in this program is as follows:

$90.00 per day for a maximum of 260 days.

Travel expenses will be reimbursed for one trip from the scholar's normal location to the Air Force facility at the start of the appointment; and one return trip from the laboratory to the scholar's normal home base at the end of the appointment period. This travel will be reimbursed in accordance with SCEEE travel policy.

An expense allowance of $35.00 per day will be reimbursed for each day the scholar spends at the Air Force Weapons Laboratory location during the research year.

III. INFORMATION BROCHURE FOR AFWL RESEARCH SCHOLARS

The information brochure was furnished to all appointed scholars. The purpose was to inform them of their responsibilities and obligations under the program and to furnish guidance in properly completing their invoices for compensation and reimbursement of expenses.

Also appended to this brochure is a copy of the Budget Memorandum to be completed by the scholar. This established the program budget against which the individual scholars could bill SCEEE for their services.

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION (SCEEE)

1984 USAF-SCEEE
WEAPONS LABORATORY RESEARCH SCHOLAR PROGRAM

INFORMATION BROCHURE

for

WEAPONS LABORATORY RESEARCHERS

April 1984

INFORMATION BROCHURE

for

AFWL SCHOLARS

Table of Contents

I. AFWL SCHOLAR OBLIGATIONS

SCEEE is required by contract to impose certain obligations on you in your status as an AFWL Scholar. This section outlines those obligations, and you should read them thoroughly. You are required to sign and return the statement of understanding before the final processing of your appointment can be completed. The following is a list of these obligations:

1. Research Goals and Objectives: A statement of research objectives must be provided to SCEEE near the beginning of the Research period. It should outline your goals and the approach you intend to follow in researching these goals. It should be submitted with your first invoice for payment. Neither travel expenses nor expense allowances will be reimbursed until after receipt of your statement of research objectives. The report should also clearly indicate the date of your first working day of the research period.

2. Final Report: At the end of your research effort, you are required to submit to SCEEE a completed, typewritten scientific report stating the objective of the research effort, the approach taken, results, and recommendations. Information on the required format is included in a "FINAL REPORT INFORMATION BULLETIN" which will be furnished to you at a later time. However, the final report must be approved by the AF Weapons Laboratory and then transmitted so as to reach SCEEE by Monday October 10, 1984. Payment of "Compensation" for the final four weeks of your research period cannot be made until SCEEE has received and approved this report in the required format.

3. Program Evaluation Questionnaire: You will be asked to complete a critique form at the end of your research period regarding your impressions of the program. This critique form should be completed and returned to SCEEE no later than Monday October 10, 1984. Return of this form is a program requirement.

4. U.S. Air Force-SCEEE-AFWL Scholar Relationship:
The U.S. Air Force and SCEEE understand and agree that the services to be delivered by the AFWL Research Scholar under this contract will be non-personal services and the parties recognize and agree that no employer-employee or master-servant relationships will exist between the U.S. Air Force and the AFWL Research Scholar. Non-personal services are defined as work performed by an individual who is responsible for an end item (such as a report), free of supervision of the U.S. Air Force and free of an employer-employee relationship.

As an AFWL Research Scholar, you will not:

(a) Be placed in a position where you are appointed or employed by a Federal Officer or are under the supervision, direction, or evaluation of a Federal Officer, military or civilian.

(b) Be placed in a staff or policy-making position.

(c) Be placed in a position of command, supervision, administration; or control over Air Force military or civilian personnel or personnel of other contractors or become a part of the U. S. Air Force organization.

The services to be performed under the AFWL Research Scholar Program do not require SCEEE or the AFWL Scholar to exercise personal exercise personal judgement and discretion on behalf of the U.S. Air Force; rather, the AFWL Scholar will act and exercise personal judgement and discretion in coordination with the Air Force Weapons Laboratory Technical Focal Point.

The Air Force will have unrestricted use of and access to all data developed during the period of this appointment.

II. ALLOWABLE TRAVEL EXPENSES

The AFWL Research Scholar Program provides potential funding for one round trip between your home and your assigned research location. As soon as you have signed and returned your appointment letter along with the budget sheet, you will be authorized to receive reimbursement for travel expenses as described below.

As outlined in the SCEEE-AFWL Research Scholar Obligations section in this brochure, you are authorized reimbursement for travel to your assigned research location at the start of your research effort and a return trip at the end of the research period. You are expected to make your own arrangements for this travel; after each trip you may invoice SCEEE for reimbursement of allowable expenses in the format described in the Instructions for Invoicing for Compensation and Reimbursement section of this brochure. Closely coordinate your travel plans with your TECHNICAL EFFORT FOCAL POINT AT THE LABORATORY. Your Technical Effort Focal Point is an individual at your research location who will be identified prior to your effort start date.

All travel reimbursements under SCEEE-AFWL Research Scholar appointments are made according to current SCEEE policy, and deviations from the approved budget are not authorized and will not be reimbursed. In light of these restrictions, you may choose either to travel by commercial airline at coach rates or less, by bus, by driving your private auto, or by a combination. (Please note that funding for rental cars is not allowed; SCEEE will not reimburse this expense. With any of these choices you may claim reimbursement up to the amount for the most direct routing, taking into the account the desirability of routing on interstate highways if you drive your private auto.

Reimbursement for direct route travel by commercial airline will thus be paid on your submission of an invoice to SCEEE following the invoicing instructions referenced above. In the view of the convenience of having a car at the research location, SCEEE strongly recommends that a private auto be used for travel when practical. Reimbursement for mileage when you drive your private auto is at the rate of 20¢ per mile within the routing restrictions mentioned above and will likewise be paid on submission of an invoice prepared according to the referenced instructions.

These items above are the only reimbursable travel allowances authorized for the AFWL RESEARCH SCHOLAR appointment. Please be advised that any additional travel expenses incurred during the appointment period will be your personal responsibility. However, travel expenses incurred at the request of the Laboratory should be arranged through your Laboratory Technical Focal Point and will be covered by other travel budgets. This other travel is not covered by the AFWL Research Scholar Program.

During the Research period, you will be authorized to receive an expense allowance in lieu of a per diem payment. The rate of this allowance is $35 per day for a maximum of 365 days. To receive this allowance, you are required to invoice for it as described in the invoicing reference above.

13

III. INSTRUCTIONS FOR INVOICING FOR COMPENSATION AND REIMBURSEMENT

Attached is a copy of the Invoice Format that you are required to use to obtain compensation or other reimbursement from SCEEE. Note that all disoursements by SCEEE for compensation, travel, and/or other expenses are subject to audit approval, so you must submit receipts substantiating charges invoiced.

In addition, you must prepare and attached to each completed invoice a <u>Brief Report of Effort.</u>

A. PREPARATION OF BRIEF REPORT OF EFFORT

Whenever you submit an Invoice for reimbursement to SCEEE you must also include a brief report describing your activities for the invoice period. To meet this obligation, <u>you must prepare, date, sign, and attach to your com-</u> <u>pleted invoice a Brief Report of Effort describing the research accomplished on the appointment and explain any travel</u> during the invoice period.

This report should include innovative techniques and designs or discoveries which may be disclosed as patents. Rights to any inventions or discoveries shall reside with SCEEE unless determined otherwise by the contracting agency.

The Brief report should never exceed one typewritten page and most often should be considerably shorter than one page.

The following is an example of such a report:

BRIEF REPORT OF EFFORT

Effort has been initiated on pole extraction methods. The modified ordinary least squares technique has been giving fair results. Work is presently being done on finding a better matrix inversion technique for the case when the coefficient matrix is ill-conditioned. Some problems have been encountered with conditioning when the data is filtered.

Travel invoice is for the trip to my research location.

_____ November 18, 1983

14

B. PREPARATION OF INVOICE FORMAT

Detailed instructions on properly completing your Invoice Format for reimbursement are provided below. Review them carefully.

(1) In the opening statement of the claim for remuneration on the invoice format, two dates are required. They are the date of your appointment letter from SCEEE (in the first blank) and the date you signed that letter accepting your appointment (in the second blank).

Other financial items required on the Invoice Format are for COMPENSATION, TRAVEL, EXPENSE ALLOWANCE. These are now explained individually with examples.

(2) COMPENSATION

(a) In the first blank to the right of COMPENSATION indicate the number of days you are claiming for compensation in this particular invoice.

(b) In the next blank enter your SCEEE-AFWL RESEARCH SCHOLAR daily appointment rate as noted in your appointment letter.

(c) Multiply the number of days times your appointment rate and enter the total dollar amount in the blank at the far right side. Note that the accumulated total number of days you claim on this appointment may not exceed the number authorized in your appointment letter. Some specific details on the compensation days must be provided in the next space.

(d) Under the heading Date, list the date of each of the days you are claiming for compensation, and opposite each date under the heading Place of Activity indicate where you worked on that date.

A sample entry of a correctly completed COMPENSATION item is shown below:

SAMPLE COMPENSATION ENTRY ON INVOICE

COMPENSATION: (12 days @ $90.00 per day)....... $ 1080.00 (II)

Date (Specify exact dates)   Place of Activity

Nov 3-4, 1983   AF Weapons Laboratory
Nov 7-11, 1983 (inclusive)   Kirtland AFB, New Mexico
Nov 14, 15, 16, 17, 18, 1983

15

(3) TRAVEL

   (a) Under the heading Date indicate the date you departed on your trip and the date you arrived at your destination.

   (b) Under the heading Departure/Arrival Time list the departure and arrival times for the corresponding days you listed under Date.

   (c) List your destination under the heading Destination.

   (d) Under the heading Mode, indicate your principal means of conveyance; i.e., commercial air, private auto, etc.

   (e) Under the heading Amount, itemize these expenditures for travel reimbursement.

   (f) Total these travel items and enter the total dollar amount to be reimbursed for travel in this particular submission on the line to the right of Total Travel Expense.

An example of a correctly completed TRAVEL entry is shown below.

TRAVEL EXAMPLE: TRAVEL TO RESEARCH LOCATION BY PRIVATE AUTO

TRAVEL: (Attach receipts for all Airline or Bus charges. Payment cannot be made without receipts attached to invoice.)

Date	Departure/Arrival Time	Destination	Mode	Amount
9/27-10/1/83	0630/1530	AF Weapons Lab. Kirtland AFB, NM.	Private Auto	$300.00

One-way trip from home in Eugene, Oregon to Kirtland AFB, NM.
         (1500 mi x 20¢/mi= $300.00)
               (mileage at start: 24162; at end: 25662)

Total Travel Expense .... $ 300.00 (III)

Please note the following comments about the TRAVEL EXAMPLE:

   i) Travel by your private auto in lieu of a commercial airline is authorized as a convenience to the traveler.

   ii) Travel with use of a privately-owned vehicle will be reimbursed at the rate of 20¢ per mile provided mileage is listed with the start and end mileage on each separate use for all distances over 100 miles.

(4) <u>EXPENSE ALLOWANCE</u>

This item on the invoice will be used to claim the $35 per day for reimbursement of costs incurred at your assigned research location.

(a) In the first blank to the right of EXPENSE ALLOWANCE enter the number of days for which you are claiming reimbursement of the expense allowance for costs incurred at your assigned research location.

(b) Multiply this number by the daily allowance rate of $35.00 and enter this total dollar amount in the blank at the far right.

(c) <u>Itemize the specific days</u> for which you are claiming the Expense <u>Allowance reimbursement.</u> It can include weekend days and holidays as well as regular work days.

The following is a sample of a correctly completed <u>EXPENSE ALLOWANCE</u> item.

<u>SAMPLE</u>

<u>EXPENSE ALLOWANCE</u>: (__16_days @ $35.00/day)........ $__560.00__ (IV)

       Specific dates covered:
       11/3/83 - 11/18/83 (inclusive)

---

(5) You may combine reimbursement requests for compensation, travel, and expense allowance in the same invoice. The total for all items invoiced should be indicated on the blank labeled "<u>GRAND TOTAL FOR INVOICE</u>" in the lower right hand side of line 5.

(6) IMPORTANT: Indicate in the space provided on <u>each</u> invoice the address to which you want the check mailed.

(7) You must sign and date your invoice in the lower right hand corner as "VENDOR" before it is submitted; you MUST also have your Laboratory Technical Focal Point <u>countersign</u> the invoice before it is mailed to SCEEE.

Invoices should be mailed to:

       AFWL RESEARCH SCHOLAR PROGAM OFFICE
       SCEEE Central Florida Facility
       1101 Massachusetts Avenue
       St. Cloud, Florida 32769

AFWL RESEARCH SCHOLAR PROGRAM
INVOICE FORMAT

(Brief Report of Effort Attached)

1. I claim remuneration from SCEEE, Inc. via the terms and conditions of the agreement
   dated _____ and accepted _____ as follows:

2. COMPENSATION: (____days @ $_____ per day)....................$_____(II)

   Date (Specify exact dates)   Place of Activity

3. TRAVEL: (Attach receipts for all Airline or Bus charges. Payment
           cannot be made without receipts attached to invoice.)

   Date   Departure/Arrival Time   Destination   Mode   Amount

   Total Travel Expense .... $_____(III)

4. EXPENSE ALLOWANCE: (_____days @ $35.00/day).................... $_____(IV)

   Specific dates covered:

5. GRAND TOTAL FOR INVOICE (Sum of II, III, IV above).............. $_____(V)

6. Please send check to following address:

7. I certify that compensation invoice is not concurrent with compensation received
   from other Federal government projects, grants, contracts, or employment.

   X_____   X_____
   LAB. TECHNICAL FOCAL POINT SIGNATURE   VENDOR SIGNATURE

   Location of Faculty_____   Social Sec. No._____

   Telephone _____   Telephone_____

   Date _____   Date _____

SOUTHEASTERN
CENTER FOR
ELECTRICAL
ENGINEERING
EDUCATION (SCEEE)

BUDGET MEMORANDUM

TO: All AFWL Research Scholars
FROM: AFWL Research Scholar Program Director
DATE: September 1, 1983
SUBJECT: 1983-1984 AFWL Research Scholar Budget

Attached is a partially completed budget for your AFWL Research Scholar Program appointment which we ask you to complete according to the following instructions and return to SCEEE. Note that the budget requires you to complete the travel item.

The travel budget has not been completed for you because you have two options on how you may complete this travel: (1) by way of commercial airline, bus or train or (2) by way of your private auto. The private auto mode of travel is strongly recommended due to the convenience of having an auto at your research location. You may request funding for one round trip between your home and your assigned research location.

The travel item authorizes reimbursement for either travel by your private auto @ 20¢/mile or travel by commercial airline (coach), bus or train. If you choose to drive your own vehicle on this trip, estimate the total round-trip mileage you will drive and enter that figure in the appropriate blank under "Travel by Private Auto". Then multiply that figure by $0.20 and enter the dollar amount in the appropriate space at the right under "Travel by Private Auto".

If you choose to travel by way of commercial airline, bus or train call your selected carrier and request the rate for round trip coach fare between your home and your assigned research location. Itemize the rate and write the total in the appropriate space beside "Travel by Common Carrier".

19

If your travel will be a mixture of common carrier and private auto, enter the estimated figures as described in both of the above cases.

Rental cars are not authorized for the research period and will not be reimbursed by SCEEE unless specific approval has been obtained from SCEEE for exceptional circumstances. Except for the mileage or travel fare, expenses incurred enroute to the research location and return for the research period are not reimbursable.

DO NOT TOTAL THE BUDGET.

After completing the above steps, please enter your name and address in the space provided in the upper left corner of the sheet, enter your research period starting date, sign and enter your social security number on the indicated blank at the bottom right corner, and return the budget sheet to SCEEE with your appointment letter.

TENTATIVE APPOINTMENT BUDGET

AIR FORCE WEAPONS LABORATORY RESEARCH SCHOLAR PROGRAM

Name _____   Research Period Starting Date _____

Address _____

_____

_____

_____

BUDGET ITEM

1. Time compensation: [260 days @ $ 90.00 per day] . . . . $ 23,400.00

2. Travel

    Travel by common carrier: [coach fare or less]. . . . .   _____

    Travel by private auto: [____ miles @ 20¢ per mile] . . . .   _____

3. Other expenses:
    Expense allowance: [365 days @ $35 per day] 12,775.00

4. TOTAL   _____

5. SCEEE G&A @ 8.3%   _____

6. Total from federal agency   _____

7. Total invoicing under your appointment may not exceed: _____

Signature _____

Social Security Number _____

21

IV. QUESTIONNAIRES AND RESPONSES

Questionnaires were sent to the Air Force Weapons Laboratory Research Scholars and their technical associates at the laboratory to obtain their evaluation of the program. The responses are furnished here.

1983-1984
AIR FORCE WEAPONS LABORATORY SCHOLAR PROGRAM
EVALUATION QUESTIONNAIRE
(TO BE COMPLETED BY PARTICIPANT)

Name _Rodney L. Williamson_

Department (at home) _Chemistry_

Home Institution _I came from Univ of Washington - I'm going to Sandia Nat'l L_

Research Colleague: _Dr. Steven J. Davis_

Laboratory Division of Colleague: _AFWL / ARDD_

Brief Title of Research Topic: _Intracavity Gain Detection_

A. TECHNICAL ASPECTS

1. Was your research assignment within your field of competency and/or interest?

   YES _✓_ NO ____ .

2. Did you have a reasonable choice of research assignment? YES _✓_ NO ____ .
   If no, why?_____

3. Was the work challenging? YES _✓_ NO ____ .
   If no, what would have made it so?_____

   _____

4. Would you classify your year activity as research? YES _✓_ NO ____ .
   Comment:_____

5. Were your relations with your research colleagues satisfactory from a
   technical point of view? YES _✓_ NO ____ .
   If no, why?_____

   _____

6. Suggestions for improvement of relationship. _none_____

   _____

   _____

7. Were you afforded adequate facilities and support? YES _✓_ NO ___ .
   If no, what did you need and why was it not provided?_____

   _____

AUG 20 1984

8. During this program calendar period of one year: Did you accomplish:
More than _____ , less than _____ , about what you expected __✓__?

9. Were you asked to present seminars on your work and/or your basic expertise?
YES __✓__ NO _____.
Please list dates, approximate attendance, length of seminars, and title of presentations. *AFWL/ARDD Division Seminar*
July 14, 1984 - A 1 hour presentation on
Intracavity Absorption & Gain Measurement

Approx. 20-25 persons in attendance

10. Were you asked to participate in regular meetings in your laboratory?
YES __✓__ NO _____.
If yes, approximately how often?_____ *weekly*

11. Did you perform travel on behalf of the laboratory? YES_____ NO __✓__.
Where to?_____
_____Purpose?_____
_____

12. Give a list of any "special" meetings you may have attended or participated in, such as conferences, visiting lectures, etc.
none

13. Other comments concerning any "extra" activities.___ *none* ___
_____
_____

14. On a scale of A to D, how would you rate this program?

	A (HIGH)..........D (LOW)			
Technically challenging	Ⓐ	B	C	D
Future research opportunity	A	Ⓑ	C	D
Professional association	A	Ⓑ	C	D
Enhancement of my academic qualifications	A	Ⓑ	C	D
Enhancement of my research qualifications	A	Ⓑ	C	D
Overall value	A	Ⓑ	C	D

B. ADMINISTRATIVE ASPECTS

1. How did you first hear of this program? _I was contacted by the weapons lab in response to a resume' I had sent in_

2. What aspect of the program was the most decisive in causing you to apply?
the nature of the research — I wanted to enhance my research qualifications in laser spectroscopy

3. Considering the time of year that you were required to accept or reject the offer, did this cause you any problems of commitment? YES___ NO ✓.
How could it be improved? _I felt that the time period between when I was informed of my acceptance into the program and the time I was to report to work was rather short._

4. After your acceptance, was information (housing, location, directions, etc.) supplied to you prior to the research period satisfactory? YES ✓ NO___.
How could it be improved?_____

_____

5. Did you have any difficulty in any domestic aspects (i.e., locate suitable housing, acceptance in community, social life, any other "off-duty" aspects)?
YES___ NO ✓.
If yes, please explain._____

_____

6. How do you rate the stipend level? Meager___ Adequate ✓ Generous___.

7. Would an expense-paid pre-program visit to the work site have been helpful?
Essential__ Convenient__ Not worth expense__.
Please add any other comments you may have. _Prior to my acceptance I was invited down by the weapons lab to give a seminar on my dissertation research and to tour the facility._

8. Please suggest names of organizations, <u>mailing lists</u>, journals, or other information you think would be helpful in announcing this program.

C & E news

9. Considering the many-faceted aspects of administration, how do you rate the overall conduct of this program?

Excellent___ Good ✓ Fair___ Poor___.

Please add any additional comments.

I felt the tax status of the appointment should have been made more clear at the outset

10. Please comment on what, in your opinion, are:
   a. Strong points of the program: _research opportunities & facilities, salary is competitive_

   b. Weak points of the program: _having to send away for and receive pay checks through the mail; not being able to attend conferences and meetings in other locations without losing salary_

11. On balance, do you feel this has been a fruitful, worthwhile, constructive professional experience? YES ✓ NO___.

12. Other remarks: _All things considered, I felt that the program offered a good opportunity extend my expertise in the chosen research field and that it has been successful. I found my association with Dr. S. J. Davis to be particularly rewarding._

THANK YOU

1983-1984
AIR FORCE WEAPONS LABORATORY SCHOLAR PROGRAM
EVALUATION QUESTIONNAIRE
(TO BE COMPLETED BY PARTICIPANT'S RESEARCH COLLEAGUE)

Name _JAMES A KEE_ , _MAJ_   Title _Chief, Technology Branch_

Division/Group _Aircraft + Missile Div_   AF Weapons Laboratory

Name of Participant _RONALD STANDLER_

A. TECHNICAL ASPECTS

1. Did you have personal knowledge of the AFWL Scholar's capabilities prior to arrival at work site?   YES _X_ NO____   If yes, where/how/what? _Interview_ _prior along with resume._

2. Was the AFWL Scholar prepared for his project?   YES _X_ NO____.

3. Please comment on his preparedness/competency/scope/depth of knowledge of subject area: _Extremely competent - at head of field_

4. Please comment on the Scholar's cooperativeness, diligency, interest, etc.
Extremely cooperative and flexible, and a hard worker

5. Did work performed by the AFWL Scholar contribute to the overall mission program of your laboratory?   YES _X_ NO____   If yes, how? _Advanced technology of methods for hardening electrical circuits/components against pulses_

6. Would you classify the effort under the AFWL Scholar Program as research?
YES _X_ NO____.
Comment: _Definitely advancing the state of the art_

JUL 20 1984
JF

7. Were your relations with the AFWL Scholar satisfactory from a technical point of view? YES _X_ NO ____ .
Suggestions as to how they might be improved: _____
_____

8. Do you think that by having this AFWL Scholar working with your group, others in the group benefited and/or were stimulated by his presence?
YES _X_ NO ____ . Comments: _Undoubtedly stimulated all_
working in area. Especially younger officers

9. Do you feel that the introduction to each other, together with the Scholar's work experience and performance could form a sound basis for continuation of effort by the Scholar at his resident institute? YES _X_ NO ____ .
If yes, how? _Grants and/or unsolicited proposals_
_____

If no, why not? _____
_____

10. One of the objectives of this program is to identify sources of basic research capability and availability to the USAF. On a scale of A to D, how effective to you think this program will be in that respect?

A (HIGH)....D (LOW)
A (B) C D

11. Also, please evaluate the role of this program as:

	A (HIGH)........D (LOW)			
a) An opportunity to stimulate group activity	(A)	B	C	D
b) A vehicle to improve professional interactions	(A)	B	C	D
c) A constructive, professional experience	(A)	B	C	D

B. ADMINISTRATIVE ASPECTS

1. When did you first hear of this program?___March 83 (Approx)___
_____

2. Were you involved in selecting, screening or prioritizing the faculty appli-
cants for your division? YES _X_ NO ____.
If yes, do you have any suggestions for improvement of the procedures used?
_____No_____

3. Would an expense-paid pre-program visit to the work site be advisable?
Essential__ Convenient _X_ Not worth expense __.
Please add any comments:__Only after initial screening__
__established suitability_____

4. Please comment on the program length. Were you as a team able to accomplish:
More than ___, less than ___, about what you expected _X_ ?
Comments:____More in some areas, less in component__
__testing due to sluggishness of A.F. procurement__
_____system.__

5. Would you desire another AFWL Scholar to be assigned to you and/or your group
/division? YES _X_ NO ___.
If no, why not?_____

6. Overall, how would you rate the administration of the program?
       A (HIGH).....D (LOW)
       (A) B C D

7. Other remarks:_____
_____
_____
_____
_____

THANK YOU

1983-1984
AIR FORCE WEAPONS LABORATORY SCHOLAR PROGRAM
EVALUATION QUESTIONNAIRE
(TO BE COMPLETED BY PARTICIPANT)

Name _____

Department (at home) _____

Home Institution _____

Research Colleague: _Major James Kee_____

Laboratory Division of Colleague: ___NTHT_____

Brief Title of Research Topic: _Transient Protection of Electronics_

A. TECHNICAL ASPECTS

1. Was your research assignment within your field of competency and/or interest?
   YES _X_ NO ____ .

2. Did you have a reasonable choice of research assignment? YES _X_ NO ____ .
   If no, why? _____

3. Was the work challenging? YES _X_ NO ____ .
   If no, what would have made it so? _____
   _____

4. Would you classify your year activity as research? YES _X_ NO ____ .
   Comment: _____

5. Were your relations with your research colleagues satisfactory from a
   technical point of view? YES _X_ NO ____ .
   If no, why? _no one in branch has expertise in my area_
   _____

6. Suggestions for improvement of relationship. _None_____
   _____
   _____

7. Were you afforded adequate facilities and support? YES ____ NO _X_ .
   If no, what did you need and why was it not provided? _____
   _Components for testing and prototypes took 3 months
   to order, and longer to receive them, laboratory
   equipment and bench space not available
   consistently_

8. During this program calendar period of one year: Did you accomplish:
    More than_____, less than_X_, about what you expected_____?

9. Were you asked to present seminars on your work and/or your basic expertise?
    YES_X_NO____.
    Please list dates, approximate attendance, length of seminars, and title of presentations.

    13 July   2 hours "transient protection techniques" ~18 peq
    Dec?   1 hour   "literature searches"   ~12 people

10. Were you asked to participate in regular meetings in your laboratory?
    YES_X_NO____.
    If yes, approximately how often? _weekly seminars_

11. Did you perform travel on behalf of the laboratory? YES_X_NO____.
    Where to?   _FT. Bliss_
    _____ Purpose? _psychiatric interview_
    for security clearance

12. Give a list of any "special" meetings you may have attended or participated in, such as conferences, visiting lectures, etc.

    None

13. Other comments concerning any "extra" activities._None_

    _____

    _____

14. On a scale of A to D, how would you rate this program?

	A (HIGH)..........D (LOW)		
Technically challenging	A	(B)	C	D
Future research opportunity	(A)	B	C	D
Professional association	A	B	(C)	D
Enhancement of my academic qualifications	A	B	C	(D)
Enhancement of my research qualifications	A	B	C	(D)
Overall value	A	(B)	C	D

B. ADMINISTRATIVE ASPECTS

1. How did you first hear of this program? _Advertisement in Spectrum_

2. What aspect of the program was the most decisive in causing you to apply?
personal contacts for research contracts

3. Considering the time of year that you were required to accept or reject the offer, did this cause you any problems of commitment? YES _X_ NO ___.
How could it be improved? _More advance planning_

4. After your acceptance, was information (housing, location, directions, etc.) supplied to you prior to the research period satisfactory? YES ___ NO _X_.
How could it be improved? _Information Nonexistent, but I didn't need it_

5. Did you have any difficulty in any domestic aspects (i.e., locate suitable housing, acceptance in community, social life, any other "off-duty" aspects)?
YES ___ NO _X_.
If yes, please explain. _no social life_

6. How do you rate the stipend level?   Meager _X_ Adequate ___ Generous ___.
My 12 month salary in academia is more than twice the AFUL salary

7. Would an expense-paid pre-program visit to the work site have been helpful?
Essential _X_ Convenient ___ Not worth expense ___.
Please add any other comments you may have. _I paid for one out of my own pocket._

8. Please suggest names of organizations, <u>mailing lists</u>, journals, or other information you think would be helpful in announcing this program.

you have good coverage now

9. Considering the many-faceted aspects of administration, how do you rate the overall conduct of this program?

Excellent___ Good___ Fair X Poor___.

Please add any additional comments.

pay invoices slow to be processed
Moving expense reimbursement needs improvement
security clearance bungled

10. Please comment on what, in your opinion, are:

a. Strong points of the program: _(1) opportunity to establish contacts for continued_
(2) use of AFWL library   _funding_

b. Weak points of the program: _lack of laboratory facilities in NTPT_

11. On balance, do you feel this has been a fruitful, worthwhile, constructive professional experience? YES X NO___.

12. Other remarks:_____

THANK YOU

Name _Dr. Charles P. Luehr_

Department (at home) _—_

Home Institution _—_

Research Colleague: _Major Raymond L. Bell_

Laboratory Division of Colleague: _Civil Engineering Research Division (AFWL/NTED)_

Brief Title of Research Topic: _Local Computational Mesh Refinement and Hydrocode Enhancements._

A. TECHNICAL ASPECTS

1. Was your research assignment within your field of competency and/or interest?

    YES _X_ NO ___ .

2. Did you have a reasonable choice of research assignment? YES _X_ NO ___ .
If no, why? _____

3. Was the work challenging? YES _X_ NO ___ .
If no, what would have made it so? _____

_____

4. Would you classify your year activity as research? YES _X_ NO ___ .
Comment: _Partly as research, but mainly as code development._

5. Were your relations with your research colleagues satisfactory from a technical point of view? YES _X_ NO ___ .
If no, why? _____

_____

6. Suggestions for improvement of relationship. _—_ _____

_____

_____

7. Were you afforded adequate facilities and support? YES _X_ NO ___ .
If no, what did you need and why was it not provided? _____

_____

8. During this program calendar period of one year: Did you accomplish:
More than ____, less than ____, about what you expected _X_?

9. Were you asked to present seminars on your work and/or your basic expertise?
YES _X_ NO ___ .
Please list dates, approximate attendance, length of seminars, and title of presentations. *A few short briefings on my work to small groups of about 2 to 5 people.*

10. Were you asked to participate in regular meetings in your laboratory?
YES _X_ NO ___ .
If yes, approximately how often? *occasionally*

11. Did you perform travel on behalf of the laboratory? YES ___ NO _X_ .
Where to? _____
_____ Purpose? _____
_____

12. Give a list of any "special" meetings you may have attended or participated in, such as conferences, visiting lectures, etc. *Attended the Annual Joint APS/AAPT (American Physical Society / American Association of Physics Teachers) Meeting 1/30/84 to 2/2/84 in San Antonio, TX. Also attended the following AAPT Workshops just before and after the meeting: "Introduction to Microcomputers as Laboratory Instruments" and "Advanced Microcomputer Interfacing."*

13. Other comments concerning any "extra" activities. _____
_____
_____

14. On a scale of A to D, how would you rate this program?

	A (HIGH).........D (LOW)			
Technically challenging	Ⓐ	B	C	D
Future research opportunity	A	Ⓑ	C	D
Professional association	A	Ⓑ	C	D
Enhancement of my academic qualifications	A	Ⓑ	C	D
Enhancement of my research qualifications	Ⓐ	B	C	D
Overall value	Ⓐ	B	C	D

B. ADMINISTRATIVE ASPECTS

1. How did you first hear of this program? _Advertisement in three issues of the "Physics Today" publication (see Physics Today, July 1953, page 108)_

2. What aspect of the program was the most decisive in causing you to apply?
   Research opportunity. Also location.

3. Considering the time of year that you were required to accept or reject the offer, did this cause you any problems of commitment? YES____ NO _X_ .
   How could it be improved? __—__

4. After your acceptance, was information (housing, location, directions, etc.) supplied to you prior to the research period satisfactory? YES _X_ NO____ .
   How could it be improved? __—__

5. Did you have any difficulty in any domestic aspects (i.e., locate suitable housing, acceptance in community, social life, any other "off-duty" aspects)?
   YES____ NO _X_ .
   If yes, please explain. __—__

6. How do you rate the stipend level? Meager____ Adequate____ Generous _X_ .

7. Would an expense-paid pre-program visit to the work site have been helpful?
   Essential____ Convenient _X_ Not worth expense____ .
   Please add any other comments you may have. _Convenient, but not essential in my case._

8. Please suggest names of organizations, <u>mailing lists</u>, journals, or other information you think would be helpful in announcing this program. *To announce it to mathematicians, I suggest* <u>*SIAM News*</u> *(newspaper pub. by the Society for Industrial and Applied Mathematics)* <u>*Notices of the American Mathematical Society*</u> *(journal pub. by the American Mathematical Society)*

9. Considering the many-faceted aspects of administration, how do you rate the overall conduct of this program?

     Excellent **X** Good___ Fair___ Poor___ .

Please add any additional comments.

I had many questions to ask. John Ungvarsky and Major Raymond Bell of AFWL, and Warren L. Steele of SCEEE were very helpful in answering my questions and making suggestions. Also many others helped too.

10. Please comment on what, in your opinion, are:

     a. Strong points of the program: *Opportunity to participate in an AFWL research program, and to become acquainted with some AFWL research efforts and some of its research staff.*

     b. Weak points of the program: ___ *none* ___

11. On balance, do you feel this has been a fruitful, worthwhile, constructive professional experience? YES **X** NO___ .

12. Other remarks: *I'm very glad to have participated.*

THANK YOU

1983-1984
AIR FORCE WEAPONS LABORATORY SCHOLAR PROGRAM
EVALUATION QUESTIONNAIRE
(TO BE COMPLETED BY PARTICIPANT'S RESEARCH COLLEAGUE)

Name _Dr. David C. Straw_ Title _Deputy Pgm Manager_

Division/Group _Advanced Technology Div_ AF Weapons Laboratory

Name of Participant _Dr. Paul Kolen_

A. TECHNICAL ASPECTS

1. Did you have personal knowledge of the AFWL Scholar's capabilities prior to arrival at work site? YES ____ NO _X_ If yes, where/how/what? _____

2. Was the AFWL Scholar prepared for his project? YES ____ NO _X_.

3. Please comment on his preparedness/competency/scope/depth of knowledge of subject area: _Dr. Kolen is an experienced experimentalist. Whereas he has limited knowledge in the specific area of work, his broad background places him in a good position to contribute._

4. Please comment on the Scholar's cooperativeness, diligency, interest, etc. _He was very cooperative in the early period — finding a mutually satisfactory research project._

5. Did work performed by the AFWL Scholar contribute to the overall mission program of your laboratory? YES ____ NO ____ If yes, how? _Too early to assess. Dr. Kolen is staying till (at least) summer '85._

6. Would you classify the effort under the AFWL Scholar Program as research? YES _X_ NO ____.

Comment: _____

7. Were your relations with the AFWL Scholar satisfactory from a technical point of view? YES _ᐱ_ NO ____ .
Suggestions as to now they might be improved: _____

_____

8. Do you think that by having this AFWL Scholar working with your group, others in the group benefited and/or were stimulated by nis presence?
YES _ᐱ_ NO ____ . Comments: _____

_____

9. Do you feel that the introduction to each other, together with the Scholar's work experience and performance could form a sound basis for continuation of effort by the Scholar at his resident institute? YES ___ NO ___ .
If yes, how? _____

_____

If no, wny not? _____

_____

10. One of the objectives of this program is to identify sources of basic research capability and availability to the USAF. On a scale of A to D, how effective to you tnink this program will be in that respect?

A (HIGH)....D (LOW)

A ↑B C D
  ↑

11. Also, please evaluate the role of this program as:

	A (HIGH)........D (LOW)			
a) An opportunity to stimulate group activity	A	B	C	D
b) A vehicle to improve professional interactions	A	B	C	D
c) A constructive, professional experience	A	B	C	D

B. ADMINISTRATIVE ASPECTS

1. When did you first hear of this program? _____1953, June_____
   _____

2. Were you involved in selecting, screening or prioritizing the faculty appli-
   cants for your division? YES _✓_ NO ____.
   If yes, do you have any suggestions for improvement of the procedures used?
   Well done! Perhaps earlier if possible,
   would be better,

3. Would an expense-paid pre-program visit to the work site be advisable?
   Essential___ Convenient _✓_ Not worth expense___.
   Please add any comments: _____
   _____

4. Please comment on the program length. Were you as a team able to accomplish:
   More than ____, less than _✓_, about what you expected ____?
   Comments: _____
   _____

5. Would you desire another AFWL Scholar to be assigned to you and/or your group
   /division? YES ____ NO ____.
   If no, why not?_____
   _____

6. Overall, how would you rate the administration of the program?

   A (HIGH).....D (LOW)

   A B C D

7. Other remarks:_____
   _____
   _____
   _____
   _____

THANK YOU

1983-1984
AIR FORCE WEAPONS LABORATORY SCHOLAR PROGRAM
EVALUATION QUESTIONNAIRE
(TO BE COMPLETED BY PARTICIPANT)

Name _Paul T. Kolen_

Department (at home) _Physics Dept_

Home Institution _Utah State U_

Research Colleague: _Dr. David Stray_

Laboratory Division of Colleague: _Air Force Weapons Lab / Electron beam Facility_

Brief Title of Research Topic: _bombardment of surfaces with energetic electrons_

A. TECHNICAL ASPECTS

1. Was your research assignment within your field of competency and/or interest?

    YES _X_ NO ___ .

2. Did you have a reasonable choice of research assignment? YES _X_ NO ___ .
   If no, why?_____

3. Was the work challenging? YES _X_ NO ___ .
   If no, what would have made it so?_____
   _____

4. Would you classify your year activity as research? YES _X_ NO ___ .
   Comment:_____

5. Were your relations with your research colleagues satisfactory from a
   technical point of view? YES ___ NO _X_ .
   If no, why? _no one in branch had any expertise in This_
   Topic

6. Suggestions for improvement of relationship. _none, simply a need_
   for more personel with appropriate backgrounds
   _____

7. Were you afforded adequate facilities and support? YES _X_ NO ___ .
   If no, what did you need and why was it not provided?_____
   _____

8. During this program calendar period of one year: Did you accomplish:
   More than _____, less than _X_, about what you expected _____?

9. Were you asked to present seminars on your work and/or your basic expertise?
   YES _X_ NO _____.
   Please list dates, approximate attendance, length of seminars, and title of
   presentations. _Aug + Oct of 1984, Two Talks about 1 hour in length_
   Approx: 15-25 at each. Title of Talk was "Large Scale
   Anisotropies in The Universe"

10. Were you asked to participate in regular meetings in your laboratory?
    YES _X_ NO _____.
    If yes, approximately how often? _once a month_

11. Did you perform travel on behalf of the laboratory? YES _X_ NO _____.
    Where to? _Los Alamos, Univ of Michigan, Oak Ridge Nat Lab,_
    Martin Merrietta Corp. Purpose? _To discuss physics appropriate_
    To research activity

12. Give a list of any "special" meetings you may have attended or participated
    in, such as conferences, visiting lectures, etc.
    None

13. Other comments concerning any "extra" activities. _____
    _____
    _____

14. On a scale of A to D, how would you rate this program?

	A (HIGH)..........D (LOW)		
Technically challenging	A (B)	C	D
Future research opportunity	(A) B	C	D
Professional association	(A) B	C	D
Enhancement of my academic qualifications	A B	(C)	D
Enhancement of my research qualifications	(A) B	C	D
Overall value	A (B)	C	D

B. ADMINISTRATIVE ASPECTS

1. How did you first hear of this program? _Advertisement in Physics Today_

2. What aspect of the program was the most decisive in causing you to apply?
research Topic

3. Considering the time of year that you were required to accept or reject the offer, did this cause you any problems of commitment? YES___ NO _X_ .
How could it be improved?_____

4. After your acceptance, was information (housing, location, directions, etc.) supplied to you prior to the research period satisfactory? YES _X_ NO___ .
How could it be improved?_____

5. Did you have any difficulty in any domestic aspects (i.e., locate suitable housing, acceptance in community, social life, any other "off-duty" aspects)?
                         YES___ NO _X_ .
If yes, please explain._____

6. How do you rate the stipend level?   Meager___ Adequate _X_ Generous___ .

7. Would an expense-paid pre-program visit to the work site have been helpful?
    Essential _X_ Convenient___ Not worth expense___ .
Please add any other comments you may have. _____

8. Please suggest names of organizations, <u>mailing lists</u>, journals, or other in-
formation you think would be helpful in announcing this program.

None come To mind That are not already included

9. Considering the many-faceted aspects of administration, how do you rate the
overall conduct of this program?

Excellent___ Good _X_ Fair___ Poor___ .

Please add any additional comments.

10. Please comment on what, in your opinion, are:
   a. Strong points of the program: _a great deal of research Freedom_
   & working hours

   b. Weak points of the program: _The incredibly long lead Time_
   required To aquire equipment & materials To accomplish research.

11. On balance, do you feel this has been a fruitful, worthwhile, constructive
professional experience? YES _X_ NO___ .

12. Other remarks:_____

THANK YOU

APPENDIX I

In this appendix we have listed some program statistics, a list of the meetings scholars attended or had co-authored presented papers, and a list of the participating scholars with some background information. The three categories within this appendix are:

A. Program statistics

B. Technical meetings or conferences attended.

C. List of 1983-1984 AFWL Research Scholars

A. PROGRAM STATISTICS

A.1 Applications received —————————————————————————————————27

A.2 Number of scholars appointed————————————————————————————04

       Highest Degrees Awarded:
       Number holding doctorate degree   4
       Number holding masters degree   0

A.3 Average age of participating scholars—————————————————————38

A.4 The following four (4) technical disciplines were
represented by the 4 initial appointments.

       Applied Mathematics
       Chemistry
       Physics
       Physics & Astronomy

A.5 The four (4) universities from which scholars received their
graduate degrees are as follows:

       University of Washington,
           Seattle, Washington

       New Mexico Institute of Mining & Technology
           Socorro, New Mexico

       University of California
           Berkeley, California

       Utah State University
           Logan, Utah

A.6 The scholars were chosen from a variety of geographical locations.
The locations of their primary residences are:

       Logan, Utah
       Corvallis, Oregon
       Webster, New York
       Seattle, Washington

B. Technical meetings or conferences attended and bibliography
of papers presented or published while in residence.

B.1 <u>Technical meetings</u>: Attendees....1

American Physical Society, San Antonio, Texas

C. <u>List of 1983-1984 AFWL Research Scholars</u>

NAME/ADDRESS	DEGREE/SPECIALTY	
Dr. Paul T. Kolen Utah State University Physics & Elec. Eng. Dept. U.M.C. 41 Logan, Utah 84322 (801) 750-2890	Degree: Specialty:	Ph.D., Physics, 1983 Particle Beam Dynamics and Guided Electromagnetic waves.
Dr. Charles P. Luehr Oregon State University Dept. of Mathematics Corvallis, Oregon 97331 (503) 754-4686	Degree: Specialty:	Ph.D., Applied Mathematics, 1962 Boundary Value Problems of Mathematical Physics.
Dr. Ronald B. Standler Rochester Inst. of Tech. Electrical Engineering Dept. One Lomb Memorial Drive Rochester, New York 14623 (716) 475-2411	Degree: Specialty:	Ph.D., Physics, 1977 Atmospheric Electricity
Dr. Rodney L. Williamson University of Washington Chemistry Department BG-10 Seattle, Washington 98195 (206) 543-2258	Degree: Specialty:	Ph.D., Physical Chemistry, 1983. Molecular Spectroscopy

APPENDIX II

1983-1984 AFWL Research Scholar Final Reports

In this appendix we have collected the statements of Goals and
Objectives provided by the scholars after their consultation with
Geophysics Laboratory technical personnel. A list of their research
report titles and a compilation of report abstracts are also included.

        A. Goals and Objectives
        B. List of Research Reports
        C. Abstracts of Research Reports.

II.A   RESEARCH GOALS AND OBJECTIVES

My interest in the AFWL Research Scholar Program is directed toward the ongoing work in particle beam generation, acceleration, and propagation. Also of interest to me is the current work being done with free electron lasers. Most of my past experience has been directed toward these aforementioned topics and it is my desire to continue in this research field. I feel that my current research, which includes relativistic kinematics and relativistic electromagnetic theory, would be very useful in the context of particle beam dynamics. As shown on the enclosed resumé, I have an extensive programming and electronics background as it pertains to experimental research.

Paul E. Kolen

CONTENTS

ACKNOWLEDGEMENT

1983-1984 AFWL-SCEEE WEAPONS LABORATORY SCHOLAR PROGRAM

Sponsored by the

AIR FORCE WEAPONS LABORATORY

Conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

LOCAL MESH REFINEMENT AND
OTHER METHODS FOR HYDROCODES

Prepared by: Dr Charles P. Luehr

Research Location: Air Force Weapons Laboratory
                        Civil Engineering Research Division
                        Technology Branch
                        Atmospheric Phenomenology Section (AFWL/NTEDA)
                        Kirtland AFB, NM 87117

AFWL Research Contact: Major Raymond L. Bell

Date: September 25, 1984

Contract Number: F49620-82-C-0035

REFERENCES

1. Halverson, H., and S. N. Bunker, "Definition Study of High Energy Accelerator Systems for Applications in Space," FINAL REPORT, SPIRE Corporation, Patriots Park, Bedford MA 01730, September 1982.

2. Ritchie, R. H., and H. B. Eldridge, "Optical Emissions from Irradiated Foils I," Phys. Rev. 126 (6), p. 1935, (1962).

3. Frank, A. L., E. T. Arakawa, and R. D. Birkhoff, "Optical Emissions from Irradiated Foils II," Phys. Rev. 126 (6), p. 1947, (1962).

4. Ritchie, R. H., "Optical Emissions from Surface Plasmons," Phys. Stat. Sol. 39. p. 297, (1970).

5. Elson, J. M., R. H. Ritchie, "Photon Interactions at a Rough Metal Surface," Phys. Rev. 4 (12), p. 4129, (1971).

Acknowledgement

VII. RECOMMENDATIONS

At this writing, all design work has been completed. The assembly of the TREF has just begun with the first e⁻ beam injection into the vacuum tank scheduled for late November 1984. This is a critical test of the electron optics to determine the total beam divergence at the target. If too much beam spreading occurs resulting in too low a current density, an external longitudinal magnetic field, B_z, must be imposed. This will require the addition of an external solenoid wound on the vacuum tank proper or some other beam focusing arrangement. Once the necessity of the beam focusing has been addressed, the TREF can be made operational once the required software has been developed. Further improvements and modifications to the facility will most probably be forthcoming once in operation.

Two additional improvements to the TREF might be to introduce an ambient plasma into the vacuum tank to better simulate the space environment and allow the e⁻ gun to be pulsed and gated. The first could be implemented by backfilling the tank with the proper gas mixture/pressure combination and generate the plasma by RF discharge. The TREF has the capability of supporting this plasma with no modification to the structure.   The e⁻ gun can be gated by computer control via an infrared detector/emitter through the plexiglas wall of the PIR. The IR detector can control or pulse the voltage applied to the e⁻ gun grid.

monochrometer, 8.0, with no light loss.

The monochrometer used for this application is manufactured by Photon Technology International Inc. model #01-001 quarter meter scanning monochrometer. A nominal 1180 lines/mm grating is used yielding a first order dispersion of 32A/mm. The grating is driven by a stepper motor which in turn is controlled by the external computer via a RS-232 interface. The monochrometer output is detected by one of two side-on Hamamatsu Photomultiplier Tube (PMT). The total spectral response of the two individual tubes, models #R106, and #R758, are such that the required range of 180nm to 1000nm is adequately covered. The spectral data together with all other relevant data is formatted and stored for later analysis by the integrated data acquisition system described in the following section.

VI. DATA ACQUISITION SYSTEM

All data acquisition and instrument control is implemented through a PSP/System 1 manufactured by the Transaic Corporation. This integrated data acquisition system is comprised of an enhanced IBM PC, 25 slot CAMAC instrument crate, CAMAC crate controller with dataway interface, and a PC to CAMAC interface card plus driver software. The PC is enhanced with dual double density floppy disks, an external 20 Mb system hard disk, and 640 Kb of system RAM. All data will be stored on the hard disk while the experiment is in progress and later down loaded to the VAX 11/750 for final analysis. The functions of the TREF will be initiated via computer/CAMAC control to limit potential operator errors and allow multiple users. Vacuum valves and vents are all pneumatically operated and controlled by an AC actuator module within the CAMAC crate. The remainder of the control and data I/O is routed through the appropriate A/D and D/A CAMAC modules and interfaces to the active devices via the required interface.

the other end of the fibre to be focused onto the slits of the scanning monochrometer for analysis. The utility of this optical arrangement is apparent by the fact that the optical collected by the QF is now restricted to a narrow acceptance cone centered on a selected viewing angle. This restricted acceptance cone data allows an extrapolation to operational detector geometries. The tolerances held in the alignment and machining of FP1 define the size and error of the acceptance cone for the optical system. The design diameter of the aperture in FP1 is given as 2.00 mm +/- .05 mm which together with a 75 mm focal length for L1 translates to a half angle acceptance cone of .764° +/- 4.8% for normal light onto L1.

The optics of figure 4 can be reconfigured into an imaging mode. This differs from the detector mode by imaging the target emissions onto the FP1 focal plane. This configuration maximizes the intensity of the optical emission at the expense of increasing the acceptance cone to 1.79° +/- 2.0%. These calculations are based on an aperture diameter for L1 of 75 mm with an object distance of 1219 mm +/- 2%.

The four viewing angles allowed within the vacuum chamber are at 175°, 60°, 45°, and 30° with respect to the target normal. Due to the proximity of the optics to the e⁻ gun structure and tank end-flange at the 175° position, the optical path must be folded 180° by mirrors M1 and M2 as seen on figure 4. These folding mirrors are not necessary at the other viewing positions due to their positioning in external vacuum extensions off the appropriate access flange.

Ultimately, the QF terminates in a two lens array which collamates the emerging light from the QF with the first lens and refocuses the light onto the entrance slits of the monochrometer with the second. This lens arrangement matches the f-number of QF, 1.2, to that of the
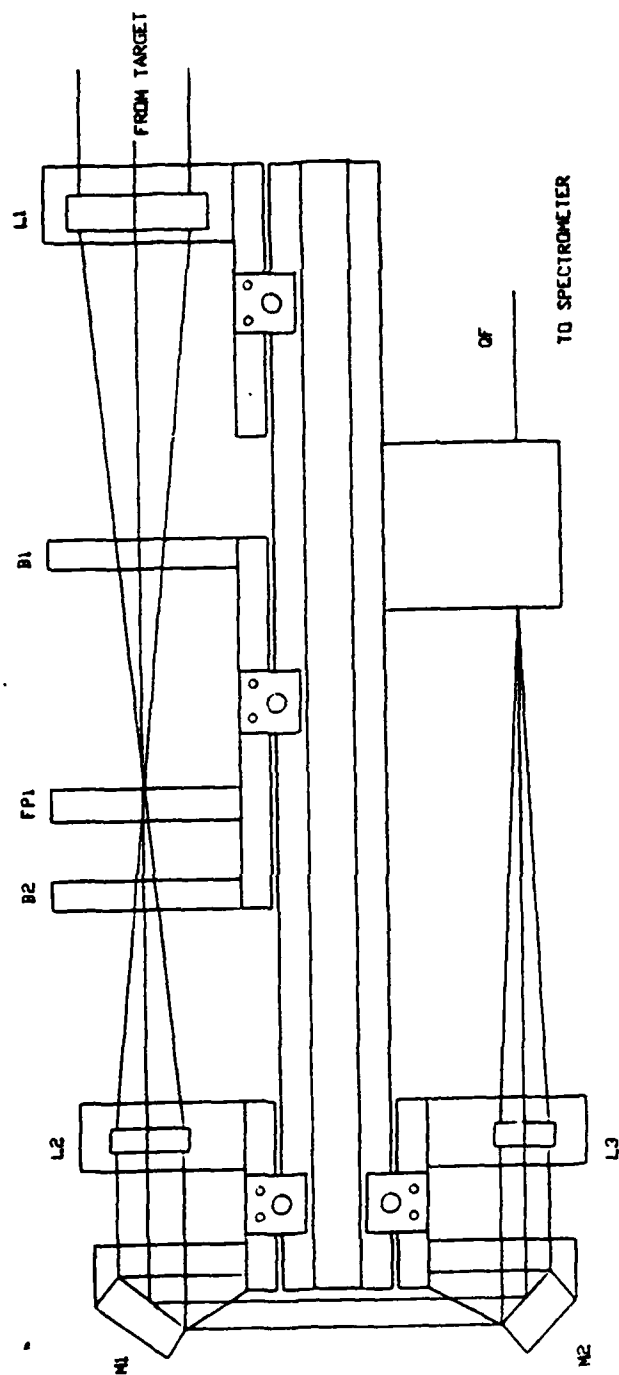
FROM TARGET

L1

B1

FP1

B2

L2

L3

M1

M2

QF

TO SPECTROMETER

FIGURE 4

plus the coolant reservoir. The electrical/thermistor/cooling lines are routed to the e⁻ gun via a SF_6 pressurized duct connecting the PIR to the nylon insert. Finally, the three power supply controls; on/off, filament current, and grid voltage are routed out of the PIR with nylon rods through pressure fittings in the plexiglas wall. These rods immediately contact a ground strap outside the wall for operator protection.

V. OPTICS

The optical subsystem for the TREF was designed to have the capability of simulating a realistic detector scenario while at the same time maintaining an ability to image within the confines of the actual experimental geometry. This scheme was chosen to require only a relative spectral calibration in lieu of an absolute calibration which is difficult to obtain even under the best of conditions. By operating the optics in the detector mode, yielding maximum angular resolution, the detector output can be scaled with respect to the specific parameters of an actual detector system. This along with the expected environmental effects can yield a reasonable estimate of the detectability of a potential spectral emission at a specific scattering angle. Configuring the optical system in the imaging mode will allow the system to operate at the highest sensitivity to study the target emissions without regard to detectability.

Figure 4 shows a structural ray trace of the optics in the detector mode. In this configuration the optics are arranged to function as a telescope, i.e. focusing light with parallel incidence to L1 from the target. Light baffles B1 and B2 straddle the focal plane of the incident parallel light, FP1, to insure that all stray light other then that focused on the FP1 plane is intercepted before arriving at L2. The light is collimated by L2 and ultimately refocused by L3 onto a 1 mm diameter quartz transmission optical fibre QF. Finally, the light emerges from
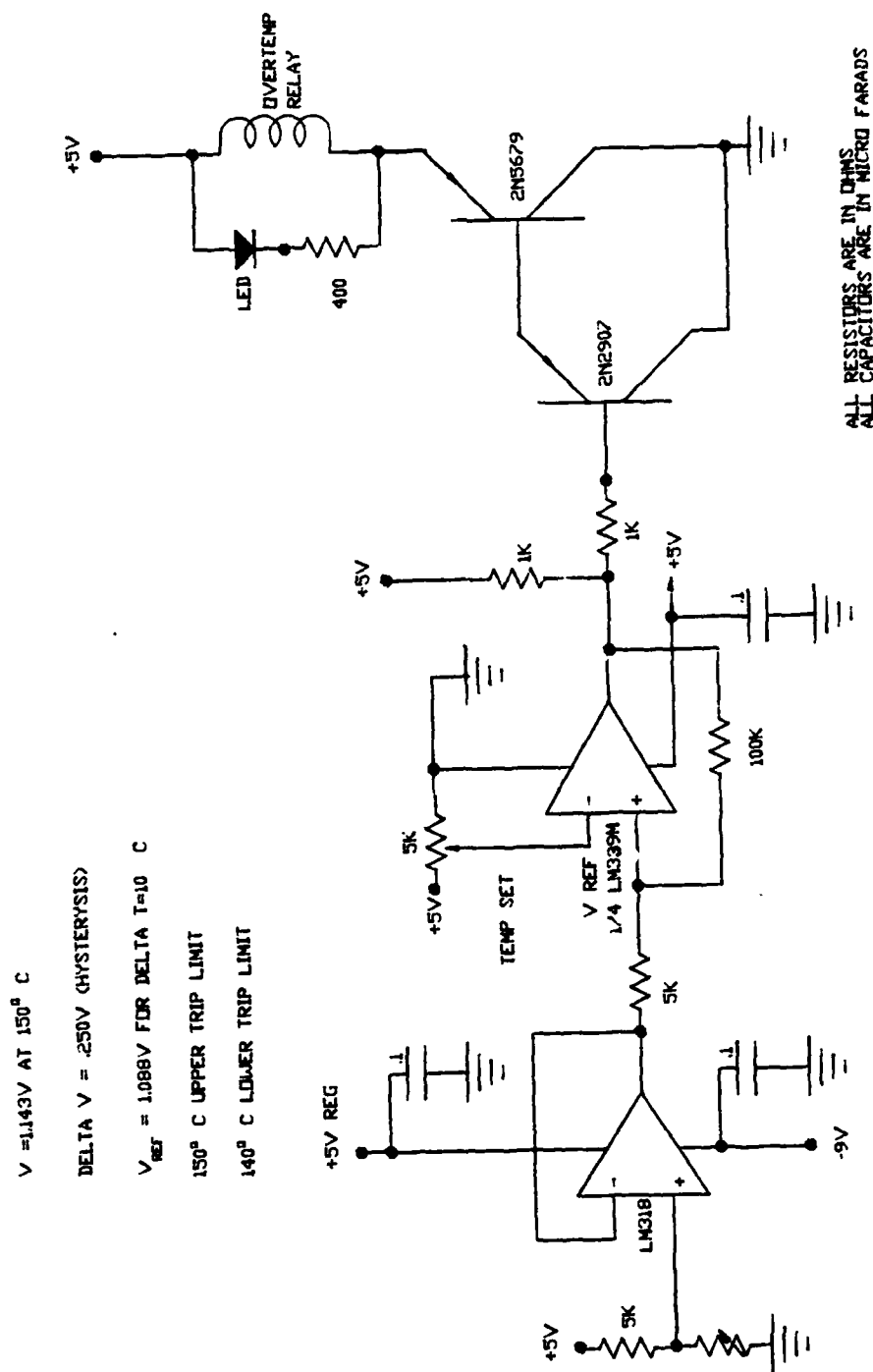
FIGURE 3

gun be operated at a floated potential with respect to the tank ground. This in turn requires that all filament and grid power supplies to the e⁻ gun be floated at the same potential. Power is coupled to these supplies from the power grid via a 100 KV, 800 KVA isolation transformer. The decision to float the gun at the accelerating potential intead of the target was made to allow the target to be biased at any arbitrary potential or floated to simulate target charging effects. Also, by biasing the gun instead of the target, emissions from secondary electron impact back onto a biased target are avoided.

The e⁻ gun filament supply is a voltage regulated 0-8v, 10 A DC supply. It incorporates a LM317T variable voltage IC regulator with an external current amplifier/limiter circuit to supply the required 8.2 A for a beam current of 20ma while limiting the maximum current to 8.5 A as recommended by the manufacturer. The grid supply is a 10% ripple, 20ma capacitively filtered DC supply. These supplies are isolated from earth ground via the isolation transformer and raised to the final accelerating potential by the external 100 KV, 100m high voltage DC power supply manufactured by Hipotronics model #8100-100.

The e⁻ gun is electrically floated within the vacuum tank by the 275mm nylon insert previously described. This nylon insert also thermally isolates the e⁻ gun from the tank requiring active thermal control and monitoring. Water is used as the coolant which is recycled through a electrically isolated reservoir. 'Figure 3 is the temperature monitor circuit which senses the e⁻ gun case temperature. A thermistor is mounted onto the case and if the temperature exceeds 150 C° the filament current is interrupted until the temperature drops below 140 C°.

For operator safety, a pressurized SF_6 plexiglas isolation rack (PIR) contains all electronics required to float at the e⁻ gun potential

FIGURE 2

atmosphere and possibly degrade the e⁻ gun filament by multiple exposures
to air. Once the target is changed, the STC is roughed down to 1 mTorr
prior to opening the gate valve and repositioning the target in the main
tank.

High vacuum of approximately 10^{-6} to 10^{-7} Torr is obtainable by the
use of a 200 mm cryo-pump in conjunction with the usual mechanical
roughing pump. The cryo-pump, manufactured by CTI-Cryogenics model type
Cryo-Torr8, was chosen to preclude the possibility of backstreaming oil
contamination common with oil diffusion pumps. Vacuum is monitored in
both the main tank and the STC by the usual thermocouple/ion gauge
combination.

As touched upon earlier, the main backscatter angle of interest is
180° with respects to beam incidence. Due to the physical size of both
the e⁻ gun insert and optical support structure, the best that can be
obtained is 175°. Along with the 175° viewing angle, three additional
100 mm ports are provided to allow viewing at 30°, 45°, and 60° with
respects to the target normal. When not in use, these ports are blanked
off to reduce tank volume. The outgasing components of the targets can
be identified once the target is repositioned within the main tank. A
computer interfaced residual gas analyzer (RGA) manufactured by Dycor
Electronics Inc. model #M100 is used for this application. The RGA has a
mass range of 1-100 amu.

IV. E-GUN POWER SUPPLY/ACCELERATOR SYSTEM

Figure 2 is a schematic of the e⁻ gun power supply/accelerator
system. The e⁻ gun is a commercial flood type gun manufactured by
Kimball Physics Inc. Model #EGG-3A. It has a nominal beam current rating
of 20 ma with a peak current of 80 ma. Internally, the gun is capable of
holding off up to 5 KV accelerating potential with a grounded case. For
this application, the accelerating potential of 100 KV requires that the

FIGURE 1

RESIDUAL GAS ANALYSER

SMALL TARGET CHAMBER

TARGET ACTUATOR

VACUUM GATE VALVE

TARGET

OPTICAL VIEWING PORTS

MAIN VACUUM TANK

VACUUM GATE VALVE

ELECTRON GUN

NYLON INSERT

OPTICAL SYSTEM

CRYO-PUMP
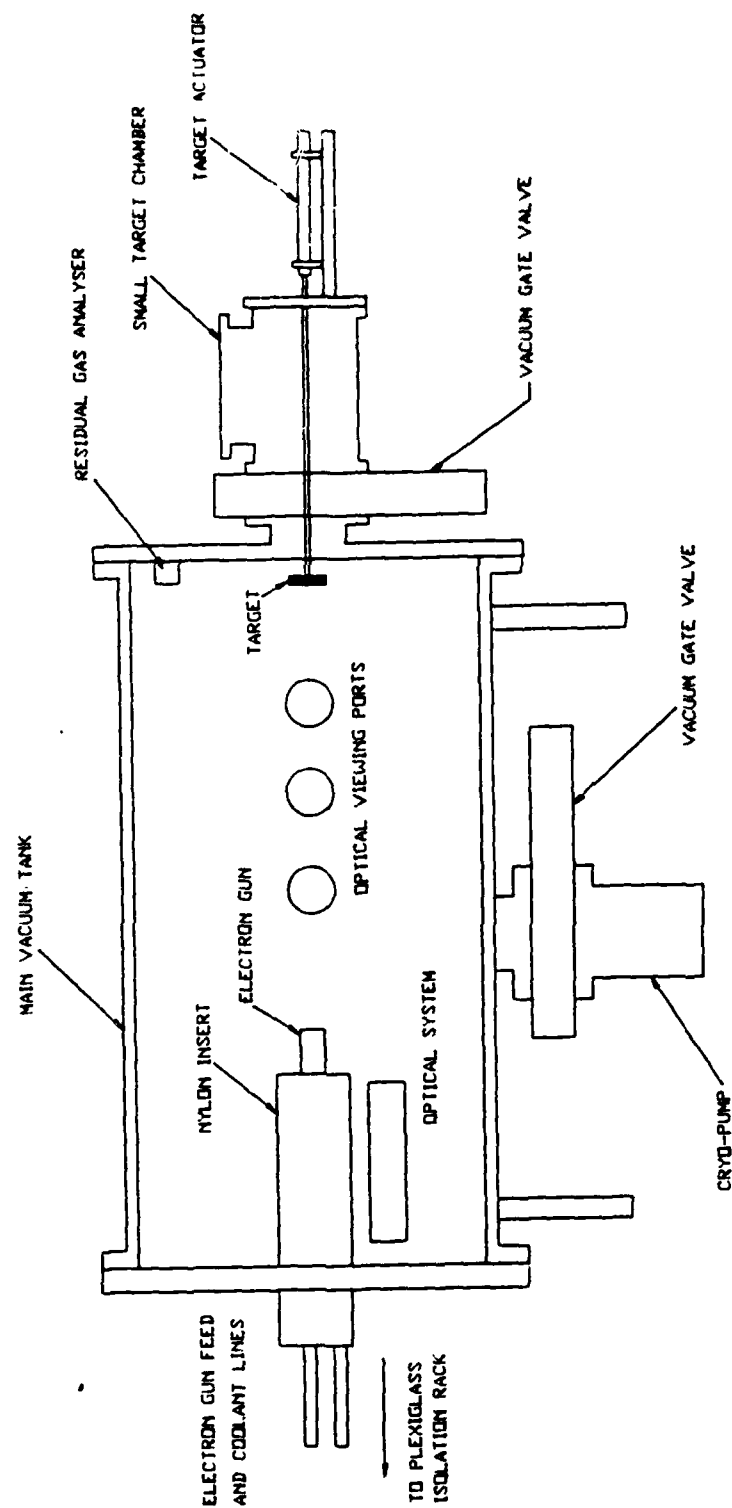
ELECTRON GUN FEED
AND COOLANT LINES

TO PLEXIGLASS
ISOLATION RACK

(3) Before and during the electron bombardment, the outgasing and boil-off constituents will be analyzed by a residual gas analyzer (RGA). The RGA will make multiple scans over the range of 1-100 amu at preselected times to obtain the time history of the outgasing elements.

(4) During electron bombardment, the monochrometer will record the time history of preselected lines or will scan the optical band. The optical signal can be observed at 175°, 60°, 45°, and 30° with respect to the target normal. Also, the polarization of the emission will be observed by a motor driven polarizer.

All the relevant data from each target will be stored on hard disk via the computer controller for later analysis.

III. VACUUM TANK

Figure 1 is a scale drawing of the TREF vacuum tank. The interior dimensions of the tank are 1524 mm long by 762 mm in diameter. These dimensions were chosen as a compromise between tank volume and the desire to collect optical data as close to 180° from beam incidence as possible. The e⁻ gun is electrically isolated from the tank ground potential via a 275 mm nylon insert on the e⁻ gun end-flange. This insert extends the working end of the e⁻ gun 300 mm into the tank yielding a net gun to target distance of 1224 mm. The e⁻ gun power supply connections and cooling lines, also isolated from ground potential, are routed to the gun through this same insert.

The target, with a maximum diameter of 100 mm, is introduced into the vacuum tank by a synchronized vacuum gate valve/pneumatic operated plunger. When the gate valve is closed, the plunger is retracted into a small target chamber (STC). The STC allows targets to be changed out and reintroduced into the beamline without having to vent the main chamber to

II. OBJECTIVES:

The project described in the rest of this report will deal with the design, construction, and operation of an experimental facility designed to study the optical emissions between 180 nm to 1000 nm generated by the interaction between a 50-100 KeV e⁻ beam and a potential target in a near earth vacuum environment. This spectral range was chosen so that conventional optical techniques and elements could be used.

In order to identify potentially useful optical emissions as target signatures, the temporal, polarization, and angular amplitude dependence of the optical emissions must be understood. Coupled with the quantum efficiency of the emissions, these aforementioned amplitude variables will determine whether or not the emission will be detectable by the particular detector at the appropriate distance and direction from the target. Of equal importance will be the determination of the source of the emission, i.e. generated by the interaction of the beam with the outgassing materials, oxide layer, or with the solid surface itself. This information will help determine if the emission can be artificially suppressed or generated as a countermeasure.

In order to quantitively answer these questions, a facility will be designed as to implement the following procedures:

(1) Each sample will be introduced into the TREF vacuum chamber from atmosphere as quickly as possible.

(2) Upon introduction into the vacuum chamber, the target will be subjected to electron bombardment of the appropriate current and energy.

Unfortunately, most of these data are not applicable to the thick targets of interest. This is mainly due to the great care taken by the various experimenters to clean and prepare the target surface prior to electron bombardment. This, of course, makes sense in the context of a basic physics experiment to eliminate as many of the surface species contaminants as possible as to not complicate the spectra.

It has been shown[5] that even a monoatomic layer of oxide on the surface of a metal target and its surface roughness greatly affects the interaction mechanisms and the corresponding spectral output. Because of this strong spectral dependence on surface conditions, it is necessary to study target materials that have undergone similar surface "conditioning" as that of the actual target. This conditioning is easily accomplished by simply storing the material in the same environment as the real target for the appropriate length of time such that the material is in dynamic equilibrium with respect to such processes as gas absorption, oxidation, and other surface contamination effects. By the same reasoning, the target must be studied under such conditions to insure that processes as outgasing and surface boil-off, which are greatly affected by the ambient environment, are realistically simulated.

In order to study the spectral emissions of the e⁻ beam/surface interaction, it is necessary to be able to observe and record these emissions as a function of time, polarization, wavelength, relative intensity, and angular distribution. This is accomplished by a combination of light gathering optical elements and an optical fibre coupled to a quarter meter scanning monochrometer. The subsystems of the Target Return Experiment Facility (TREF) will be discussed individually in the following sections.

I. INTRODUCTION:

The main focus of this research is directed towards neutral atomic beam/target phenomenology, in particular, the problem of determining target status through stimulated surface emissions.

The most likely candidate[1] for the neutral beam is neutral hydrogen (H^O). The interaction of the H^O beam with a potential target can be broken down into two subsets, namely, volume interactions of the high kinetic energy protons (H^+) within the target and surface interactions of the co-moving 50-100 KeV electrons (e^-). There is no significant volume interaction associated with the electrons due to their being stripped from the H^O within the first few atomic layers of the target because of atomic collision creating a H^+ within the target.

Due to the difficulty in obtaining a 200 MeV H^O beam, no accelerator facility is currently set up with a dedicated beamline, the study will commence by using a 200 Mev H^+ beam to investigate the volume interactions and a separate 100 Kev e^- beam for the surface phenomena. As a first cut, the superposition of the effects seen in the separate beam interactions is a good approximation to what might be expected from the H^O, albeit not exact. Potentially, certain processes inherent to charged particles, i.e. transition radiation, may be present that would be suppressed by the neutral H^O and vice versa. It is hoped that these specific mechanisms can be identified and accounted for in order to obtain a more realistic representation of the H^O through the H^+ and e^- results. Ultimately, these composite results will be compared to an H^O beam when it becomes available at Brookhaven National Laboratory in late FY 85.

There exists a large body of data detailing the emissions of electron surface interactions [2], [3], [4] in the open literature.

1983-1984 AFWL-SCEEE WEAPONS LABORATORY SCHOLAR PROGRAM

Sponsored by the

AIR FORCE WEAPONS LABORATORY

Conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

DESIGN OF A 100KV ELECTRON BOMBARDMENT

FACILITY FOR SURFACE OPTICAL EMISSIONS STUDIES

Prepared by: Dr. Paul Ted Kolen

Academic Rank: Staff Physicist

Department and   Department of Physics

  University: Utah State University

Research Location: Air Force Weapons Laboratory, Advanced Technology

                    Division

USAF Research

    Contact: Dr. David C. Straw

Date: November 1, 1984

Contract No.:   F49 620-82-C-0035

APPENDIX III

1983-1984 AFWL Research Scholar Final Reports

    This appendix contains the completed final reports prepared and sub-
mitted by the AFWL Research Scholars. These reports have been accepted
by the Weapons Laboratory and approved for inclusion in this Program
Technical Report.

INTRACAVITY ABSORPTION AND GAIN DETECTION

by

Dr. Rodney L. Williamson

ABSTRACT

An intracavity gain detection apparatus was set up and tested on optically pumped systems. Both pulsed and CW gain was detected on various transitions of I_2 and IF. From the data collected on these two systems the apparatus is shown to be capable of detecting net gains as small as 0.0001. Efforts were made to detect gain on chemically pumped I_2 and IF using this technique. The pumping agent was $O_2(^1\Delta)$. The preliminary results indicate that sufficient B-state I_2 is produced in the flow, prior to dissociation to I and I^*, to give rise to measurable gain. The results on chemically pumped IF indicated insufficient number density due to the slowness of the I_2 plus F_2 reaction. Suggestions for further study are made.

Transient Protection of Electronic Circuits

by

Ronald B. Standler

ABSTRACT

Electromagnetic pulses from nuclear weapons, lighting, and electrostatic discharge are three sources of electrical overstress. Such overstress can cause failure, permanent degradation, or temporary malfunction (upset) of electronic devices and systems. This problem and general solutions are briefly reviewed. Non-linear components and circuits for protection from electrical overstress are then discussed in detail. This work emphasizes spark gaps, metal oxide varistors, and avalanche diodes. However, other components, such as semiconductor diodes, thyristors, resistors, inductors, and optoisolators are also discussed. Applications of these non-linear components are discussed in the context of signal lines, AC power lines, and DC power supplies. The final chapter discusses specific upset protection circuits.

LOCAL MESH REFINEMENT AND
OTHER METHODS FOR HYDROCODES

by

Charles P. Luehr

ABSTRACT

To develop improvements for finite difference hydrocodes, three matters were pursued. An algorithm and a Fortran code were developed for generating a locally refined computational mesh satisfying many specifications and having an arbitrary number of fine mesh zones. A short method to calculate the center of mass of the fluid part of a two dimensional partial island cell was formulated and coded. The conserved moment quantity was briefly investigated for possible applications in the design of finite difference methods for fluid dynamics that conserve this quantity.

DESIGN OF A 100KV ELECTRON BOMBARDMENT

FACILITY FOR SURFACE OPTICAL EMISSION STUDIES

by

Paul Ted Kolen

ABSTRACT

An experimental facility for the study of surface emissions
generated by 100KV electrons impacting upon surfaces is described. The
utility of these stimulated surface emissions is described within the
context of their detectability at great distances once generated by an
incident neutral beam of the appropriate energy and current density.
The individual subsystems of the Target Return Experiment Facility
(TREF) are described in detail with a discussion of the physical
contraints driving the design. Detailed or block diagrams of the sub-
systems, where appropriate, are included. Finally, suggestions for
improvements and modifications of the design once operational are
offered.

II.C   ABSTRACTS OF AFWL RESEARCH SCHOLAR RESEARCH REPORTS

II.B LIST OF RESEARCH REPORTS
1983-84 USAF-SCEEE AFWL RESEARCH SCHOLAR PROGRAM

TITLE	RESEARCH ASSOCIATE
1. Design of a 100KV Electron Bombardment Facility for Surface Optical Emissions Studies	Dr. Paul T. Kolen
2. Local Mesh Refinement and Other Methods for Hydrocodes	Dr. Charles P. Luehr
3. Transient protection of Electronic Circuits	Dr. Ronald B. Standler
4. Intracavity Absorption and Gain Detection	Dr. Rodney L. Williamson

PROPOSED RESEARCH AT AFWL BY Rodney L. Williamson

It is the objective of this research effort to develop a sensitive optical gain measurement technique to be used on candidate visible laser systems. There are several systems that are currently under investigation, including the interhalogens such as IF.

Both intracavity dye laser enhancement and photon lifetime measurement techniques will be considered. The first method involves introducing the medium on which the gain measurement is to be made into the cavity of a CW dye laser. Because the dye laser is ultrasensitive to intracavity gains, if these are present the laser output will "lock" onto the chemically inverted medium at the fluorescence wavelength. This provides the basis of a <u>very</u> sensitive technique to measure small gains in various laser media. One of the advantages of this technique is that the output frequency of the dye laser does not have to perfectly match the fluorescence transition frequency of the foreign medium. This is important since for existing extracavity techniques the pump source frequency must exactly match the frequency of the transition for which the gain measurement is being made.

The second method is an extracavity method. It involves measuring the legnth of time which an optical cavity will sustain oscillation, i.e., the photon lifetime. If the absolute loss characteristics of this cavity are known, then the only other parameter which determines the photon lifetime of the cavity is the gain of the medium it contains. Thus, by measuring the photon lifetime the gain may be uniquely determined. A convenient method for making this measurement is the cavity phase shift method. This method relies on the fact that a photon entering the cavity will be delayed a certain period of time before it exits through the output coupler. Thus, the output of this cavity due to an amplitude modulated input beam will suffer a phase shift which is proportional to the photon lifetime of the cavity. This phase shift may be measured using conventional techniques.

PROPOSED RESEARCH AT AFWL BY R. STANDLER

My time will be spent on the following tasks that are related to lightning.

Electrical breakdown of soil is relevant to the study of EMP propagation and effects on missile silos and other underground structures.

Suspending a wire from a balloon line to a rocket to the ground and firing a wire-trailing rocket backwards from a F106.

Consider various ways to trigger lightning and develop some experiments which could be conducted at Langmuir Laboratory in 1984.

Consider various ways to photograph lightning and determine the current from the density of the negative.

PROPOSED RESEARCH AT AFWL BY CHARLES P. LUEHR

1. To participate in some research tasks of interest to AFWL.

2. To act as a mathematics consultant for any scientists and engineers at AFWL who would like assistance on mathematical problems in their research.

3. _Extended_ Twistor Spaces. Twistor space, a generalization of the Dirac spinor space, serves as a representation space for the conformal group, analogously to the way the Dirac spinor space is a representation space for the Lorentz group. Dirac spinors have significant applications in theories with invariance under the Lorentz group such as relativistic quantum mechanics. Likewise, twistors have applications in theories with invariance under the conformal group and the Poincaré subgroup. One of these applications is in the formulation of a gauge action principle of gravitation based on a true spinorial type of variational principle

4. Extended twistor spaces are generalizations of twistor space. The simplest one of these is a representation space for the super-conformal group and the super-Poincaré group, which play a central role in the supergravity theories at the frontier of present day research in general relativity theory. It is proposed that the appropriate extended spaces together with their structure be explicitly defined by means of axioms, that the properties of these spaces be investigated, and that applications of these spaces be made in the formulation of supergravity theories.

CONTENTS

I. INTRODUCTION

Various finite difference numerical techniques are used in computational fluid dynamics to solve the equations of fluid flow. However, there is no general method that performs satisfactorily for all problems because of the diverse and complex features possible in fluid flow. For the best performance on a given type of problem, a technique must be specially designed to handle the particular features of that type of problem.

One kind of problem is the computational modelling of the flow resulting from a nuclear explosion in the atmosphere. Of particular interest is the motion of the shock wave as it propagates away from the point of burst and interacts with the earth's surface and solid structures on the earth. The discontinuity in the flow variables at the shock front gives rise to difficulties in the computational simulation of such problems.

Finite difference techniques are better suited for smooth flow problems due to the need to represent the state of the fluid on a discrete arrangement of cells comprising the computational mesh. When steep gradients or discontinuities occur, the computational simulation of the flow problem will be less accurate. A calculated shock front will be spread out over a distance of a few cell sizes.

In many cases, a fine enough computational mesh to reduce the spreading of shock waves will give reasonable accuracy in the calculation of such problems. A calculated shock front will then be spread out into a layer of small thickness over which the flow variables make a rapid change in their values.

As the fineness of the mesh is increased to increase the accuracy of the calculation, the computer time needed for the calculation is also increased. Thus, this approach leads to a trade-off between the accuracy and the speed of the calculation. However, one can attempt to gain both advantages by another approach, the employment of a locally refined mesh. This kind of mesh is made finer in the vicinity of shock fronts and other regions needing higher accuracy in the calculation, and is made coarser everywhere else. As the shock wave moves, its location is detected by numerical tests, and the mesh is occasionally rezoned as needed in order to always place a fine mesh in the vicinity of the shock.

The local mesh refinement method has been found to be very useful. This report describes a computer subroutine that sets up a locally refined mesh of a very general form having an arbitrary number of fine mesh zones.

Another important question is how to incorporate the boundary conditions into the differencing method. Reflective boundary conditions come from the presence of the surface of the earth and solid structures on the earth. The partial island method accounts for these boundary conditions. Partial island cells are equivalent fluid cells in the computational mesh, and they are located along the solid boundaries of the flow region. The flow variables in a partial island cell are chosen to produce the same effect on the fluid flow as the effect of the reflective boundary condition. This report describes a subroutine that calculates the center of mass of the fluid part of a two dimensional partial island cell. This calculation is a small part of a procedure used in the partial island method.

Another question concerns conserved quantities. Most differencing schemes for fluid dynamics are designed to conserve mass, momentum, and energy. Another quantity, the conserved moment quantity, might not be conserved by the differencing scheme. This report briefly considers this quantity in differencing schemes and suggests that schemes designed to also include the conservation of this quantity might improve the accuracy of the calculated flow.

II. OBJECTIVES

The research objective was generally stated as the development of various methods to help in the calculation and display of computer solutions of the equations of fluid dynamics. Under this objective, the main item to be pursued was to develop improvements in numerical methods in order to increase the accuracy and speed of calculation. Most of the effort went toward this main item.

Two other items, included in the plans in the event that time permitted, were not followed up. These were to study possible applications of differential geometry to fluid mechanics, and to design ways of graphically displaying the results of computer calculations.

The primary contributions to the research objective consisted of the completion of two specific computer programming tasks and a brief theoretical investigation.

The first task was to develop a subroutine in Fortran that sets up a locally refined finite difference mesh satisfying a certain set of requirements. This subroutine can be called by a hydrocode in order to rezone the mesh whenever needed in order to have fine mesh regions placed wherever greater accuracy is wanted, such as the vicinity of shock fronts. The subroutine rezones in one coordinate, and it can be called once for each coordinate whenever the mesh is to be rezoned.

The second task was to develop a short subroutine in Fortran that computes the center of mass of the fluid part of a two dimensional partial island cell.

The third task was a brief investigation of the conserved moment quantity in finite difference schemes.

III. LOCAL MESH REFINEMENT PROBLEM

This problem was to develop a Fortran subroutine that sets up a finite difference mesh $al = x(0) < x(1) < x(2) < \ldots < x(imx) = bpl$ on an interval (al, bpl) in one coordinate x such that a certain set of requirements on the mesh are satisfied. The first requirement is that the mesh be finer in the vicinity of each of some selected points $xf(i)$ for $i=1,2,\ldots,$ nfs, where $xf(i) < xf(i+1)$. This requirement and most of the other requirements concern the cells of the mesh and their sizes, where the i th cell is the interval $(x(i-1), x(i))$ and the cell size of the i th cell is $dx(i) = x(i)-x(i-1)$ for $i=1,2,\ldots,$ imx. Finer mesh means smaller cell sizes.

The requirements on the mesh are as follows:

Requirement 1. The following are all prescribed in advance: (a) the total number imx of cells, (b) the endpoints $al = x(0)$ and $bpl = x(imx)$ of the mesh, (c) the lower limit dxmn for the smallest cell size in the mesh, (d) the maximum allowable ratio $rajmx = 1.1$ of adjacent cell sizes with the larger divided by the smaller, (e) the upper limit $1.21*dxmn$ for the smallest cell size in the mesh.

Requirement 2. The interval (al,bpl) is divided up into fine mesh regions and coarse mesh regions, with the fine mesh regions located such that all the points $xf(i)$ are contained in the fine mesh regions. The number of fine mesh regions and their sizes are no greater than needed to include these points and also to satisfy the next requirement. The remaining parts of the interval (al, bpl) are covered by coarse mesh regions.

Requirement 3. All fine mesh regions have the same constant fine mesh cell size dxf, where dxf is to be determined during the calculation and must satisfy the inequality dxmn $\leq$ dxf $\leq$ 1.21*dxmn. Each point xf(i) is in a fine mesh region with at least 2.5 fine mesh cells between the point and the nearest coarse mesh region.

Requirement 4. Cell sizes in the coarse mesh regions depend on two parameters raj and rexp. The parameter raj, the ratio of adjacent coarse mesh cell sizes, is determined during the calculation. It is the same for all coarse mesh regions, and the inequality 1.0 < raj $\leq$ 1.1 must be satisfied. The parameter rexp, the ratio of the smallest cell size in a coarse mesh region to the fine mesh cell size dxf, is determined during the calculation. It may be different for different coarse mesh regions, and must satisfy the inequality 1.0 $\leq$ rexp $\leq$ raj.

Requirement 5. There are three kinds of coarse mesh regions. There is a coarse mesh left region at the left end of the interval (al, bpl) if there is enough room for it. There is a coarse mesh right region at the right end of the interval if there is enough room for it. There is a coarse mesh middle region between each pair of adjacent points xf(i) and xf(i+1) if there is enough room for it. These three kinds of coarse mesh regions have the following properties: (a) In a coarse mesh left region, the cell sizes are decreasing from left to right by the constant ratio raj between adjacent cells. The last cell (at the right end) has the size rexp*dxf, where rexp is determined during the calculation. (b) In a coarse mesh right region, the cell sizes are increasing from left to right by the ratio raj between adjacent cells. The first cell (at the left end) has the size rexp*dxf,

where rexp is determined during the calculation. (c) In a coarse mesh middle region, the cell sizes are increasing from left to right by the ratio raj between adjacent cells in the left half of the region, and are decreasing by the ratio raj in the right half of the region. The first and last cells (at the left and right ends) have the same size rexp*dxf where rexp is determined during the calculation.

Requirement 6. The ratio raj of adjacent coarse mesh cell sizes is to be as large as possible, consistent with all the other requirements. The fine mesh cell size dxf is to be made as small as possible, consistent with all the other requirements.

Requirement 7. The subroutine is to use information from an input mesh $xl(i)$ for $i = 0,1,2,\ldots,$ imx to determine the interval (al, bpl) and dxmn. The endpoints of this interval are taken to be al = $xl(0)$ and bpl = $xl(imx)$. The minimum cell size dxlmn of the input mesh is determined, and dxmn = dxlmn/1.1 is taken to be the value for the minimum allowable cell size.

In the original statement of the problem, the ratio raj was allowed to be different for different coarse mesh regions. However, the algorithm which was found gave the same value for raj in all coarse mesh regions. This property of the values for raj is included in the present statement of the problem as an extra requirement.

IV. ALGORITHM FOR LOCAL MESH REFINEMENT

Two approaches in the search for an algorithm will now be considered. The first approach is the most straightforward, but was abandoned because of many difficulties. However, a brief discussion of it is useful in order to reveal the complex nature of the problem. The second approach depends on an idea that simplified the search.

1. FIRST APPROACH TO AN ALGORITHM

In the first approach, let us look for an algorithm consisting of the following sequence of steps. Each step is the placement of one cell into the interval (a1, bp1) such that the cell does not overlap with cells previously placed into the interval. After the last step, the interval should be completely filled with cells lying end to end, forming a mesh that satisfies all requirements. One might try a scheme in which each cell is placed adjacent to the previous cell to be placed, starting at the left end of the interval and finishing up at the right end. Or, one might try other schemes in which a cell is not always placed adjacent to the previous cell to be placed, but placed in some other unfilled part of the interval.

With each step of the process comes two decisions: where to place the cell and what size to make the cell. But how to make these two decisions in each step is a very difficult question. For each decision, one needs to look ahead to the result obtained from that decision and all other decisions in order to be sure that the requirements on the mesh will be *satisfied*. The various combinations of choices lead to a multitude of avenues to be searched in order to find a combination that works. Also, in the process, a reasonable choice for dxf and raj must be made.

Two of the most persistent difficulties that appear here and also in the second approach are the following. The first one is that the number of cells obtained in the resulting mesh is not likely to be exactly the required number imx. The second one is that one or more unfilled gaps are almost certain to be left inside the interval. A gap might have a size incompatible with the size of an adjacent cell such that the gap cannot be made into a cell without violating the requirements on the ratio of adjacent cell sizes.

Due to these and other difficulties and the need to look ahead before each decision it becomes evident that an iterative process may be needed. One might select trial values for dxf and raj, followed by some process of arranging cells into a mesh that satisfies as many of the requirements as possible. Then this process is repeated over and over with new values for dxf and raj selected each time in an attempt to improve on the results obtained from the previous selection. Finally it would be hoped that after a number of iterations of the process, a mesh satisfying all the requirements will be obtained.

But the details for the iterative procedure is still very elusive at this point of the investigation. However, the complex nature of the problem has been revealed. What is needed is some device that simplifies the search for an algorithm.

2. SECOND APPROACH TO AN ALGORITHM

A simplifying device was found, and is the basis for the second approach in the search for an algorithm. Note that in the first approach, the basic object with which one was dealing was the cell. The device is to deal with another kind of basic object, which will be called a segment. Segments are made up of cells in convenient arrangements satisfying simple requirements.

When segments are put together to fill the interval (al, bpl), the result is also an arrangement of cells forming a mesh on the interval.

The requirements on the arrangement of cells in any segment is simple enough so that an algorithm is easily found to construct a segment. Each segment constructed will occupy a subinterval of the original interval. A segment will be specified by giving a tentative subinterval (a, bp), but the resulting segment (a, b) will often not occupy the entire tentative subinterval due to an unavoidable gap (b, bp) at the right end, which cannot be included without violating the requirments on the arrangement of cells in a segment.

Futhermore, the problem of arranging segments in the original interval (al, bpl) is simpler than the problem of arranging individual cells in the original interval. Thus we have two basic steps: arranging cells to form segments, and arranging segments to fill the original interval (al,bpl).

However, there will still be the problem of getting the wrong number of cells in the mesh and the persistent occurrence of a gap. But these problems can be solved. To get the right number of cells, an iterative procedure is used. To solve the problem of a gap, the procedure is designed so that the gap will appear at the right end of the original interval. Furthermore, the result of the iterative procedure will be an arrangement of cells such that the mesh will still satisfy all the requirements even if all cells have to be expanded slightly by the same factor in order to close the gap to fill the entire interval.

Details of the algorithm is given in the self documentation with comment statements contained in the source code listings in Appendix A.

The code consists of Subroutine MESH to carry out the main procedures, and a collection of other subroutines to do various subtasks. Subroutine SEGS arranges segments together in the original interval (al, bpl) to fill a subinterval (al, bl), usually leaving a gap (bl, bpl). With trial values for dxf and raj, Subroutine MESH repeatedly calls Subroutine SEGS, changing the value for dxf or raj each time until the right number imx of cells for a mesh is obtained. It then eliminates the gap, if any, by expanding all cells by a factor. Subroutine SEGS calls on Subroutines LRSEG, MSEG, FSEG and related subroutines, which arrange cells into segments of various kinds. Details of the purpose and method of Subroutines MESH and SEGS, and all the other subroutines are given in the self documented source code listings in Appendix A.

An important question about the performance of the code concerns Step 5 (Iteration for raj) and Step 6 of the Method for Subroutine MESH given in the listing. Under normal conditions, does there exist cases for which the number ncl of cells takes on values in the loop both higher and lower than imx, but never imx exactly no matter how many times the loop in Step 5 is gone through? If this can happen, then the code would fail to get ncl exactly equal to imx, and an error return would be made when nt exceeds ntmx. Thus, although a mesh satisfying all the conditions would exist, the code would fail to find it.

Consider the possibility that there is a value for raj which gives ncl = imx-1 but the next change in ncl is 2 regardless of how small the decrements in raj are made to be, i.e., there is a slightly lower value for raj where ncl makes a jump of 2 from ncl = imx-1 to ncl = imx+1 all at once as raj decreases. If this possibility exists, then the loop would be repeated until nt exceeds ntmx.

A possible mechanism by which ncl might jump by 2 instead of 1 is related to the fact that in coarse mesh middle and right segments, a small change in raj can in some cases produce a sudden jump in the positions of cell boundaries in that segment and in all segments to its right. Since the effect of a jump originating in a middle segment can be propagated to all other segments on its right, the jump might trigger another jump in another middle or right segment on its right. The combined effect of the two or more simultaneous jumps might possibly produce a large enough change in the arrangement of cells to make ncl jump by 2 instead of 1.

To limit the effect of a jump originating in a middle segment, the extra term -(bp-b) has been put in the expression for bp for a middle segment on line 25 in the listing for Subroutine SEGS. This term should prevent the propagation of the effect of a jump beyond the middle segment in which it originates.

It has not been proved for sure that this problem of jumps by 2 in ncl really exists. If it does, then the term -(bp-b) should eliminate it, or at least greatly reduce the probability of it occurrence. A careful and complete analysis of this problem of jumps in the mesh ("meshquakes") has not been made.

An example of the speed of calculation for subroutine MESH on the CRAY 1 computer is given by several problems with imx = 50 and nfs = 2, all of which ran in about 1/2 millisecond.

V. PARTIAL ISLAND CENTER OF MASS CALCULATION

This problem was to develop a short Fortran subroutine that computes the center of mass of the fluid part of a two dimensional partial island cell. A partial island cell is a square which is divided by a straight line into two nontrivial parts. One part is the island part. The other part is the fluid part, representing a fluid of constant density. The results can also be made applicable for a rectangular partial island cell, by using cartesian coordinates parallel to the sides with separate systems of units normalized to make the height and base equal to one in the corresponding units.

The partial island cell is defined in terms of four cell descriptors, one for each side of the square. The descriptors indicate the part of the boundary of the square that boarders on the fluid part and the part that boarders on the island part of the cell.

The method of calculation has been simplified and shortened by using the observation that the fluid part is a 3, 4, or 5 sided polygon which can be obtained by removing a right triangle from a rectangle. In some cases the triangle can be a trivial triangle with two of its vertices coinciding. The area and center of mass of the triangle and rectangle are calculated, and used in the calculation of the center of mass of the polygonal fluid part of the cell.

A challenging subtask was to find a short procedure for making use of the cell descriptors to determine the location of all the corners of the triangle and two opposite corners of the rectangle to be used in the calculation.
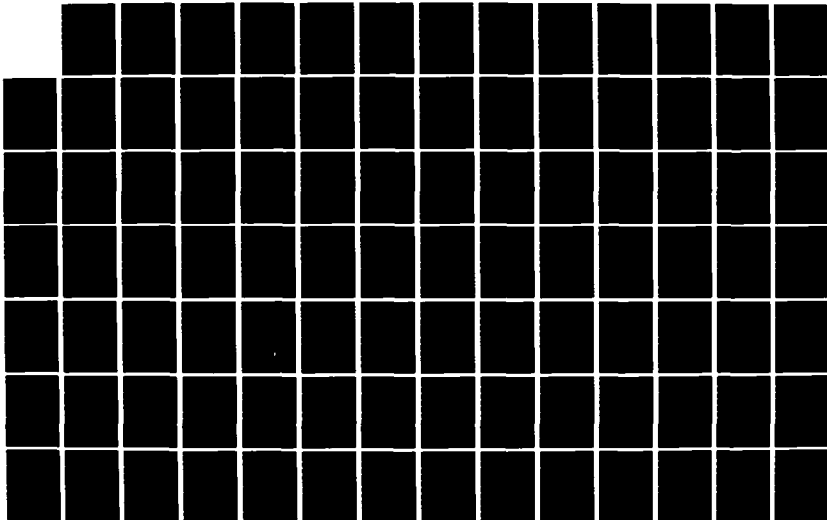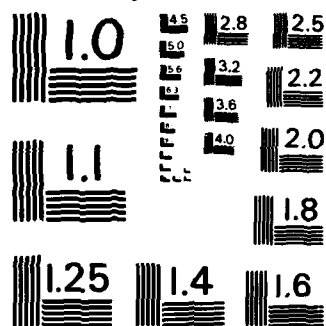
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

A detailed description of the problem and the method of calculation is given in the self-documentation with comment statements contained in the source code listing in Appendix B for the Subroutine CMASS, which carries out this calculation.

VI. THE CONSERVED MOMENT QUANTITY

To introduce the quantity to be investigated, consider a system con-
sisting of one particle of mass m in motion in 3 dimensional space. Let
$\vec{r}(t)$ and $\vec{u}(t)$ be the position vector and velocity respectively of the par-
ticle at time t. Define the moment quantity $\vec{M}(t)$ for the particle as

$$\vec{M} = m[\vec{r} - (t-t_o)\vec{u}] \tag{1}$$

where t_o is an arbitrary constant. The first term $m\vec{r}$ is the mass moment
about the origin, and the second term is $-(t-t_o)$ times the momentum $m\vec{u}$.
In the case of a free particle, $\vec{u}$ is constant, and it follows that

$$d\vec{M}(t)/dt = \vec{0} \tag{2}$$

Thus $\vec{M}$ is a constant of motion for one free particle.

For a system of n particles in 3 dimensional space, the moment quantity
$\vec{M_i}(t)$ of particle number i is

$$\vec{M_i} = m_i[\vec{r_i} - (t-t_o)\vec{u_i}], \tag{3}$$

and the total moment quantity for the system is

$$\vec{M} = \Sigma_i \vec{M_i} = \Sigma_i m_i[\vec{r_i} - (t-t_o)\vec{u_i}], \tag{4}$$

where m_i, $\vec{r_i}(t)$, and $\vec{u_i}(t)$ are the mass, position vector, and velocity
respectively of the i th particle for i = 1,2,...,n. Assume that there are
no external forces on the particles, and that the particles interact among
themselves by conservative forces. Then the total momentum is constant,
i.e.,

$$d(\Sigma_i m_i \vec{u_i})/dt = \vec{0}$$

and it follows that

$$d\vec{M}(t)/dt = \vec{0} \tag{5}$$

Thus $\vec{M}(t)$ is a constant of motion for this system.

For fluid mechanics, the total moment quantity $\vec{M}(t)$ is an integral over space,

$$\vec{M} = \int \rho[\vec{r} - (t-t_o)\vec{u}]dV. \qquad (6)$$

The part $\int \rho \vec{r} dV$ is the total mass moment about the origin, and the remaining part is $-(t-t_o)$ times the total momentum $\int \rho \vec{u} dV$. In the special case of one dimensional flow in a tube of unit cross sectional area, the total moment quantity $M(t)$ is

$$M = \int \rho[x - (t-t_o)u]dx \qquad (7)$$

where the integral is from $-\infty$ to $+\infty$.

1. BASIC EQUATIONS IN ONE DIMENSION

An example of an application of the moment quantity to finite difference schemes in fluid dynamics will be presented here. In the case considered, it will be a conserved quantity. For simplicity, select a one dimensional case governed by the following three differential equations in flux divergent form, expressing the conservation of mass, momentum and energy respectively:

$$\partial \rho / \partial t + \partial (\rho u)/\partial x = 0 \qquad (8a)$$

$$\partial (\rho u)/\partial t + \partial(\rho u^2 + P)/\partial x = 0 \qquad (8b)$$

$$\partial(\rho E)/\partial t + \partial[(\rho E + P)u]/\partial x = 0 \qquad (8c)$$

with P given by the equation of state

$$P = P(\rho, I) \qquad (9)$$

where

x = position (cm)

t = time (sec)

$\rho(x,t)$ = fluid density (g/cm^3)

$u(x,t)$ = fluid velocity (cm/sec)

$E(x,t)$ = total specific energy (ergs/g)

$I(x,t)$ = internal specific energy (ergs/g)

$P(x,t)$ = hydrostatic pressure (dyn/cm^2)

and

$$E = I + \tfrac{1}{2}u^2. \qquad (10)$$

Define the specific moment quantity $\mu(x,t)$ as

$$\mu = x - (t-t_0)u \qquad (11)$$

where t_0 is an arbitrary constant. From Equations (8a-b) and (11), the following differential equation for μ is easily derived:

$$\partial(\rho\mu)/\partial t + \partial[\rho u \mu - (t-t_0)P]/\partial x = 0 \qquad (12)$$

Equation (12), together with Equations (8), (9), and (10), can be used for deriving a finite difference scheme that conserves the moment quantity as well as the mass, momentum, and energy. It turns out, however, that the mass

moment, a non conserved quantity, can be used in place of the moment quantity
to derive the same results except for an extra term of second order in the
time step Δt. Define the specific mass moment $\lambda(x,t) = \lambda(x)$ as

$$\lambda = x. \tag{13}$$

From equations (8a-b) and (13), the following differential equation for λ
is easily derived:

$$\partial(\rho\lambda)/\partial t + \partial(\rho u \lambda)/\partial x - \rho u = 0 \tag{14}$$

2. APPLICATION TO A FINITE DIFFERENCE SCHEME

To express the differential equations in finite difference form, a com-
putational mesh is used with the x, t coordinates divided into intervals
Δx_i and time steps Δt^n respectively. The i th cell has its left and right
endpoints a_i and b_i respectively given as

$$a_i = x_{i-1/2} = x_o + \sum_{k=1}^{i-1}\Delta x_k$$

$$b_i = x_{i+1/2} = x_o + \sum_{k=1}^{i}\Delta x_k = a_i + \Delta x_i.$$

Time values $t = t^n$ for the mesh are

$$t^n = t_o + \sum_{k=1}^{n}\Delta t^k$$

where $\Delta t^k = \Delta t$ for all k in the case where all the time steps Δt^k are taken to
be equal. For each hydrovariable $f(x,t)$, where f is ρ, u, I, E or P,
denote the values at $x = x_{i-1/2}$, $x = x_{i+1/2}$, $t = t$, and $t = t^n$ as

$$f_{i-1/2} = f(x_{i-1/2}, t)$$

$$f_{i+1/2} = f(x_{i+1/2}, t)$$

$$f^n_{i-1/2} = f(x_{i-1/2}, t^n)$$

$$f^n_{i+1/2} = f(x_{i+1/2}, t^n).$$

At an arbitrary time t for cell number i, define

total mass in i th cell:

$$m_i = \int_{a_i}^{b_i} \rho \, dx \tag{15a}$$

total momentum in i th cell:

$$p_i = \int_{a_i}^{b_i} \rho u \, dx \tag{15b}$$

total energy in ith cell:

$$Q_i = \int_{a_i}^{b_i} \rho E \, dx \tag{15c}$$

total mass moment in i th cell:

$$L_i = \int_{a_i}^{b_i} \rho \lambda \, dx \tag{15d}$$

total moment quantity in i th cell

$$M_i = \int_{a_i}^{b_i} \rho \mu \, dx \tag{15e}$$

At time $t = t^n$ these quantities are denoted with a superscript n, i.e., m_i^n

etc. It is easily shown that

$$M_i = L_i - (t-t_o)p_i. \qquad (16)$$

At time t for cell number i, define the following averages:
average fluid density in i th cell:

$$\rho_i = m_i/\Delta x_i \qquad (17a)$$

average fluid velocity in i th cell:

$$u_i = p_i/m_i \qquad (17b)$$

average total specific energy in i th cell:

$$E_i = Q_i/m_i \qquad (17c)$$

position of center of mass of i th cell:

$$\lambda_i = L_i/m_i \qquad (17d)$$

average moment quantity in i th cell:

$$\mu_i = M_i/m_i. \qquad (17e)$$

At time $t = t^n$, these quantities are denoted with a superscipt n, i.e., ρ_i^n
etc. It is easily shown that

$$\mu_i = \lambda_i - (t-t_o)u_i. \qquad (18)$$

It follows that

$$m_i = \rho_i \Delta x_i \qquad (19a)$$

$$P_i = m_i u_i \qquad (19b)$$

$$Q_i = m_i E_i \qquad (19c)$$

$$L_i = m_i \lambda_i \qquad (19d)$$

$$M_i = m_i \mu_i \qquad (19e)$$

The differential equations for the quantites m_i, P_i, Q_i, L_i, and M_i are obtained from integrating Equations (8), (14), and (12) with respect to x from a_i to b_i. The result is

$$dm_i/dt = - (\rho u)_{i+1/2} + (\rho u)_{i-1/2} \qquad (20a)$$

$$dP_i/dt = - (\rho u^2 + P)_{i+1/2} + (\rho u^2 + P)_{i-1/2} \qquad (20b)$$

$$dQ_i/dt = - [(\rho E + P)u]_{i+1/2} + [(\rho E + P)u]_{i-1/2} \qquad (20c)$$

$$dL_i/dt = - (\rho u \lambda)_{i+1/2} + (\rho u \lambda)_{i-1/2} + P_i \qquad (20d)$$

$$dM_i/dt = - [\rho u \mu - (t-t_o)P]_{i+1/2} + [\rho u \mu - (t-t_o)P]_{i-1/2} \quad (20e)$$

Up to this point in the discussion, both the conserved moment quantity and the mass moment were kept. However, only one of these is needed. The mass moment will be used in the remainder of the discussion because the resulting equations are simpler.

The finite difference form of the Equations (20 a-d) will be written. For the quantities F_i, where F_i is m_i, p_i, Q_i, or L_i, its time derivative dF_i/dt at $t=t^n$ is approximated by $\Delta F_i^n/\Delta t$ where

$$\Delta t^n = \Delta t$$

$$\Delta F_i^n = F_i^{n+1} - F_i^n.$$

Then at $t=t^n$, the result is

$$\Delta m_i^n = - \Delta t [(\rho u)_{i+1/2}^n - (\rho u)_{i-1/2}^n] \quad (21a)$$

$$\Delta p_i^n = - \Delta t [(\rho u^2 + P)_{i+1/2}^n - (\rho u^2 + P)_{i-1/2}^n] \quad (21b)$$

$$\Delta Q_i^n = - \Delta t \{[(\rho E + P)u]_{i+1/2}^n - [(\rho E + P)u]_{i-1/2}^n\} \quad (21c)$$

$$\Delta L_i^n = - \Delta t [(\rho u \lambda)_{i+1/2}^n - (\rho u \lambda)_{i-1/2}^n - p_i^n] \quad (21d)$$

The values advanced to time t^{n+1} are then

$$m_i^{n+1} = m_i^n + \Delta m_i^n \quad (22a)$$

$$p_i^{n+1} = p_i^n + \Delta p_i^n \quad (22b)$$

$$Q_i^{n+1} = Q_i^n + \Delta Q_i^n \tag{22c}$$

$$L_i^{n+1} = L_i^n + \Delta L_i^n \tag{22d}$$

The right sides of the Equations (21) have quantities evaluated at the end-points of the cells. Approximations to these values can be obtained by interpolation from the corresponding average values in the two cells having the common endpoint. For the purpose of the present discussion, assume that some expressions for interpolation are given in the following form:

$$\rho_{i-1/2}^n = Z_1(\rho_{i-1}^n, \rho_i^n, \lambda_{i-1}^n, \lambda_i^n, \ldots) \tag{23a}$$

$$u_{i-1/2}^n = Z_2(u_{i-1}^n, u_i^n, \lambda_{i-1}^n, \lambda_i^n, \ldots) \tag{23b}$$

$$E_{i-1/2}^n = Z_3(E_{i-1}^n, E_i^n, \lambda_{i-1}^n, \lambda_i^n, \ldots) \tag{23c}$$

The left sides are the values at time t^n at the left endpoint $x_{i-1/2}$ of cell i. The first two arguments on the right sides are average values at time t^n in the cell i-1 and i respectively. With i replaced by i+1 in these equations, the left sides are the values at time t^n at the right endpoint $x_{i+1/2}$ of cell i.

The center of mass of the two cells having the common endpoint are included as arguments, indicating that there may be weighting factors dependent on the centers of mass. In the interpolation expressions for endpoint values of one hydrovariable, the average values of some of the other hydrovariables might also possibly appear as arguments on the right side, as indicated by the "..." notation.

Another possibility is a scheme in which the expression for each endpoint value of one hydrovariable at time t^n has that endpoint value of the same hydro-

variable at the previous time t^{n-1} as an argument on the right side. This possibility would include schemes in which endpoint values of hydrovariables are obtained at time t^n from its value at time t^{n-1} by a differencing scheme. That is, there would be two differencing steps, one for average values in the cells, and another for the endpoint values.

Finally, a differencing scheme can be described. For time t^n, assume that ρ_i^n, u_i^n, E_i^n, λ_i^n, are known. To update those variables to time t^{n+1}, the following steps are taken.

Step 1. Using Equations (19a-d) with $t=t^n$, calculate:

$$m_i^n = \rho_i^n \Delta x_i$$

$$p_i^n = m_i^n u_i^n$$

$$Q_i^n = m_i^n E_i^n$$

$$L_i^n = m_i^n \lambda_i^n$$

Step 2. Using Equations (23), interpolate to find $\rho_{i-1/2}^n$, $u_{i-1/2}^n$, $E_{i-1/2}^n$.

Step 3. Using Equations (10) and (9), calculate

$$I_{i-1/2}^n = E_{i-1/2}^n - \tfrac{1}{2}(u_{i-1/2}^n)^2$$

$$P_{i-1/2}^n = P(\rho_{i-1/2}^n, \ I_{i-1/2}^n)$$

Step 4. Using Equations (21),

calculate $\Delta m_i^n, \Delta p_i^n, \ \Delta Q_i^n, \ \Delta L_i^n$

Step 5. Using Equations (22),

calculate m_i^{n+1}, p_i^{n+1}, Q_i^{n+1}, L_i^{n+1}.

Step 6. Using Equations (17a-d) with $t = t^{n+1}$, calculate:

$$\rho_i^{n+1} = m_i^{n+1} / \Delta x_i$$

$$u_i^{n+1} = p_i^{n+1} / m_i^{n+1}$$

$$E_i^{n+1} = Q_i^{n+1} / m_i^{n+1}$$

$$\lambda_i^{n+1} = L_i^{n+1} / m_i^{n+1}$$

This differencing scheme carries the extra quantity λ_i, not considered in schemes designed only to conserve mass, momentum, and energy. The extra quantity λ_i, the location of the center of mass of the i th cell, comes about because this scheme is designed to also conserve the moment quantity. It is suggested that the inclusion of this extra quantity λ_i should increase accuracy, because λ_i contains extra information on the state of the fluid system.

VII. RECOMMENDATIONS

Subroutine MESH is a general purpose code for generating a locally refined computational mesh. Nevertheless, some potentially useful options are not now present in the code. In the following, some suggestions are given for new options that could be added to the code if needed. (a) Option to choose the fine mesh cell size dxf exactly, whenever it can be done without violating any of the other requirements on the mesh. (b) Options to relax some of the other requirements on the mesh, if necessary, in order to allow dxf to be chosen exactly. (c) Option to allow dxf to be different in different fine mesh regions. (d) Option to allow the ratio raj to be chosen exactly. (e) Option to allow raj to be chosen differently in different coarse mesh regions. (f) Option to allow regions of constant large cell size, larger than all cells in neighboring variable cell size regions. (g) Option to specify a pair of points to be covered by a single fine mesh region of smallest possible length. Also, option to specify several pairs of points to generate several fine mesh regions. Also, similar options for constant large cell size regions. (h) Option in which the endpoints of fine mesh regions can be chosen exactly, whenever it can be done without violating any of the other requirements. Also, a similar option for constant large cell size regions. (i) Options to relax some of the other requirements on the mesh, if necessary, in order to allow the end-points of fine mesh regions to be chosen exactly, when properly chosen. Also, similar options for constant large cell size regions. (j) Option to allow regions of constant intermediate cell size, larger than all the cells in the neighboring variable cell size region on one side, but smaller than

all the cells in the neighboring variable cell size region on the other side.

Further investigation of the conserved moment quantity is recomended. It is suggested that a code be written to try out the proposed one dimensional Eulerian finite difference scheme for fluid dynamics that conserves this quantity. The accuracy of this scheme could then be compared with other schemes which are not designed to conserve this quantity.

One essential part of such an investigation would concern the form of the expressions indicated in Equation (23) for the values of ρ, u, and E at the endpoints of the cells. Simple expressions representing straight-forward interpolation schemes would be easy to find. However, the best prospects might be in more elaborate expressions depending also on the values of ρ, u, and E at the endpoints from the previous time step. It would be desirable to find expressions that give good results for shock waves.

If the investigation of the one dimensional case yields good results, then a generalization of the method to the 2 and 3 dimensional cases would be good directions to continue the investigation.

In addition to Eulerian finite difference schemes, other schemes could also be modified to conserve the moment quantity. A Lagrangian scheme is an example. Another example is the scheme used in the HULL code, which is effectively Eulerian and consists of a Lagrangian step followed by advection.

A "user friendly" option for the HULL hydrocode is suggested. The purpose is to help the user to understand and to run the code by interacting with him through the CRT display and the keyboard.

It could operate as follows. First, it displays a preliminary explanation which starts with an outline of the main types of runs that can be made with the code, and what is accomplished in each. It continues by outlining the organization of the code with descriptive material and flow charts showing the flow of the control and data. The directions of flow are indicated by arrows along fixed paths, and also by the motion of symbols along the paths of flow to give a moving picture of the sequence of steps in the control and data flow. A description of the options and the input that will be expected from the user in each type of run is also included in this outline and is indicated in the flow charts. Provisions are made to omit parts of the preliminary explanations that an experienced user may choose to skip. It then outlines in more detail what the code is going to do and what the user will have to do.

After the preliminary explanations, it leads the user step by step through the procedure, and helps him to take the needed actions to run the code. It asks the user for input to be typed in, and also asks for menu selection of options from the user. Also during each part of the run, it provides information to the user such as what the code is now doing, how long it should take, and where it is in the flow chart. Finally it helps the user to decide the form of the output, and helps him to take the actions to obtain the output.

BIBLIOGRAPHY

Anderson, Robert L., and Ibragimov, Nail H., Lie-Backlund Transformations in Applications, SIAM, Philadelphia, 1979. See page 78 for the center of mass theorem, which involves the conserved moment quantity.

Bell, R. L., and Westmoreland, C., Elastic/Plastic HULL (EPHULL) Operation on the CRAY Time Sharing System (CTSS), AFWL-TR-83-6, Air Force Weapons Laboratory, Kirtland AFB, NM, April 1983.

Chambers, B. S. III, and Wortman, J. D., Two-Dimensional Shore (Partial Island) Cells for BRL HULL, ARBRL-CR-00497, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, December 1982.

Connell, Robert I., and Bell, Raymond L., User's Guide to Vector HULL on the CRAY Time Sharing System, Air Force Weapons Laboratory, Kirtland AFB, NM, to be published.

Gaby, L. P., HULL System Guide, AFWL-TR-78-115, Air Force Weapons Laboratory, Kirtland AFB, NM, January 1979.

Matuska, D. A., and Durrett, R. E., The HULL Code, A Finite Difference Solution to the Equations of Continuum Mechanics, AFATL-TR-78-125, Air Force Armament Laboratory, Eglin AFB, FL, November 1978.

Matuska, D. A., Durrett, R. E., and Osborn, J. J., HULL User Guide for Three-Dimensional Linking with EPIC3, ARBRL-CR-00484, Ballistic Research Laboratory, Aberdeen Proving Ground, MD, July 1982.

APPENDIX A

FORTRAN LISTING OF SUBROUTINES MESH,
SEGS, AND AUXILIARY SUBROUTINES

1.   subroutine mesh (imx,x1,dx1,xc1,nfs,xf,x,dx,xc)
```
c
c      SUBROUTINE MESH                                          C. F. Luehr
c
c      PURPOSE:
c
c      This subroutine sets up a finite difference mesh in one
c      coordinate.  The output mesh is finer in a neighborhood of an
c      arbitrary number of selected points, and satisfies some restrictions
c      on the ratios of adjacent cell sizes.  Also, other requirements are
c      satisfied.  It is intended to be used with a hydrocode such as the
c      HULL code, which numerically solves the equations of fluid dynamics
c      by an Eulerian finite difference method.
c      Using information from a one dimensional input mesh x1(i),
c      (i=0,....,imx), and a set of points xf(i), (i=1,....,nfs), to be
c      covered by fine mesh neighborhoods in the output mesh, it calculates
c      a one dimensional output mesh x(i), (i=0,...,imx), which partitions
c      the interval (x1(0),x1(imx)) into cells by means of partitions
c      located at x1(0)=x(0) < x(1) < x(2) < ... < x(imx)=x1(imx).  In a
c      neighborhood of each of the input points xf(i), (i=1,...,nfs), the
c      cells have the fine mesh cell size dxf, where dxf is determined
c      during the calculation of the mesh.  Everywhere else, the cell sizes
c      are as large (almost) as possible, consistent with the restriction
c      that adjacent cell sizes (the larger divided by the smaller) are
c      always between 1.0 and raj, where the ratio raj of adjacent coarse
c      mesh cell sizes is determined during the calculation of the mesh.
c      It is required that dxmn =< dxf =< 1.21*dxmn and 1.0 < raj =< 1.1,
c      where dxmn=dx1mn/1.1 and dx1mn is the smallest input cell size.
c      The value for raj is chosen to have nearly the largest possible
c      value consistent with these requirements.  It is also required that
c      the output mesh have the same number of cells imx as the input mesh,
c      and have the same endpoints (x(0),x(imx))=(x1(0),x1(imx)) as the
c      input mesh.
c      The requirements (1) through (8) listed under PURPOSE of SEGS in
c      the listing for Subroutine SEGS are satisfied, but with the
c      following changes.  Requirement (1) is replaced by the stronger
c      requirement b1=bp1.  Also, (a1,b1)=(a1,bp1)=(x1(0),x1(imx)).  In
c      Requirement (8), "3.5 fine mesh cells" is replaced by "2.5 fine
c      mesh cells," and the part in parentheses is inapplicable since
c      b1=bp1.
c      In the above and in the following discussion, " a =< b " means
c      "a less than or equal to b", and " a >= b " means "a greater than
c      or equal to b".
c
c      INPUT ARGUMENTS:
c
c      imx -- number of cells in the input mesh and in the output mesh.
c      x1(i), (i=0,....,imx) -- input mesh which partitions the interval
c      (x1(0),x1(imx)) into cells.
c      dx1(i), (i=1,....,imx) -- cell sizes for the mesh x1(i), where
c      dx1(i)=x1(i)-x1(i-1).
c      xc1(i), (i=1,....,imx) -- location of cell centers for the
c      mesh x1(i).
c      nfs -- number of points in the array xf.
c      xf(i), (i=1,...,nfs) -- points in the interval to be covered by
c      fine mesh neighborhoods in the output mesh.  It is required that
c      x1(0) =< xf(1) < xf(2) < ... < xf(nfs) =< x1(imx).
c
```

```
c     OUTPUT ARGUMENTS:
c
c       x(i), (i=0,...,imx) -- output mesh which partitions the interval
c     (x1(0),x1(imx)) into cells.
c       dx(i), (i=1,...,imx) -- cell sizes for the mesh x(i), where
c     dx(i)=x(i)-x(i-1).
c       xc(i), (i=1,...,imx) -- location of cell centers for the
c     mesh x(i).
c
c     METHOD:
c
c       1.  If the cell sizes dx1(i) of the input mesh are all equal, it
c     makes the output mesh x(i) the same as the input mesh x1(i), the
c     output cell sizes and cell centers are made the same as the
c     corresponding input quantities, and a return is made.
c       Otherwise, it goes on to the next step.
c       2.  It calculates the minimum allowable cell size dxmn=dx1mn/1.1
c     for the output mesh, where dx1mn is the minimum cell size in the
c     input mesh.  It fixes the interval for the output mesh to be
c     (a1,bp1) = (x1(0),x1(imx)).  It sets the initial value for the
c     ratio raj of adjacent coarse mesh cell sizes to rajmx=1.1, its
c     maximum allowable value.  It sets the initial value of the fine
c     mesh cell size dxf to its minimum allowable value dxmn.
c       3.  Iteration for dxf:  Using the interval (a1,bp1) as input, it
c     repeatedly calls Subroutine SEGS, which calculates b1 and the number
c     nc1 of cells in a mesh for the subinterval (a1,b1), but after each
c     call of the subroutine it increases the value of dxf before the
c     next call.  This procedure is continued until nc1 =< imx, then it
c     goes on to the next step.  But if dxf becomes too large, ie.
c     dxf > 1.21*dxmn, the calculation is terminated and an error return
c     is made.
c       4.  If nc1=imx, it goes to Step 7 (Mesh Calculation).  Otherwise,
c     it calculates the initial value of the ratio rrj < 1.0 of successive
c     trial values for raj, and gives the initial value 0 to the number
c     nt of times through the next loop.
c       5.  Iteration for raj:  Using (a1,bp1) as the input interval,
c     it repeatedly calls Subroutine SEGS, which calculates b1 and the
c     number n1 of cells in a mesh for the subinterval (a1,b1), but
c     before each call it multiplies raj by the factor rrj and increments
c     nt; the repetition is continued until nc1 >= imx, then it goes to
c     the next step.  But if raj becomes too small, ie. raj =< 1.0, or if
c     nt becomes too big, ie. nt >= ntmx, then an error return is made.
c       6.  If nc1=imx, it goes to Step 7 (Mesh Calculation).  If
c     nc1 > imx, meaning that it overshot its goal of reaching nc1=imx
c     exactly, then it divides raj by rrj to give raj its previous value,
c     calculates a larger value for rrj closer to 1.0 so that successive
c     values of raj will be closer together, and it goes back to Step 5
c     (Iteration for raj).
c       7.  Mesh Calculation:  Having found values for dxf and raj that
c     gives nc1=imx exactly, it calls the Entry XSEGS into Subroutine SEGS
c     to calculate ib1, b1, and the mesh x(i), (i=ia1,...,ib1) [where
c     ia1=0, and ib1=imx ], for the subinterval (a1,b1).  It checks the
c     value of ib1, and if it is not equal to imx, an error return is
c     made.  Otherwise, it goes on to the next step.
c       8.  It calculates the expansion factor rb needed for the expansion
c     of all cell sizes in the mesh on (a1,b1) in order to fill the the
c     entire interval (a1,bp1) exactly.  If the expanded fine mesh cell
c     size rb*dxf exceeds its maximum allowable value 1.21*dxmn, an error
```

```
      c    return is made.  It expands the mesh x(i), (i=ial,i+1), to fill the
      c    entire interval (al,bpl).  It calculates cell sizes and cell
      c    centers, and returns.
      c       Before the expansion, the mesh x(i) on the interval (al,bl)
      c    satisfied the requirements (1) through (8) listed under PURPOSE of
      c    SEGS in the listing for Subroutine SEGS.  After the expansion of
      c    all cell sizes, the new expanded interval (al,bpl) coincides with
      c    the entire interval (al,bpl), and the resulting mesh x(i) on the
      c    interval (al,bl)=(al,bpl) still satisfies the requirements (1)
      c    through (8), except that in requirement (8), "3.5 fine mesh cells"
      c    is replaced by "2.5 fine mesh cells".
      c
      cXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
      c
  2.         dimension xl(1),dxl(1),xcl(1),xf(1),x(1),dx(1),xc(1)
  3.         parameter (rmns=1.1, rajmx=1.1, ial=0)
  4.         parameter (clxf=3.5, crxf=2.5)
  5.         parameter (rdxf=1.01, rdxmn=1.21)
  6.         parameter (frrjmx=0.02, rfrrj=0.2, ntmx=50)
      c
      c    Step 1:
      c    If cell sizes dxl(i) of input mesh xl(i) are all equal, then make
      c    output mesh x(i) the same as the input mesh xl(i), make the output
      c    cell sizes and cell centers same as the corresponding input
      c    quantities, and return.
      c       Otherwise go to Step 2.
      c
  7.         dxlmn = dxl(1)
  8.         i = 2
  9. 10      continue
 10.         if(i.gt.imx) go to 20
 11.         if(dxlmn.ne.dxl(i)) go to 30
 12.         i = i+1
 13.         go to 10
 14. 20      continue
 15.         x(0) = xl(0)
 16.         do 25, i=1,imx
 17.         x(i) = xl(i)
 18.         dx(i) = dxl(i)
 19.         xc(i) = xcl(i)
 20. 25      continue
 21.         return
      c
      c    Step 2:
      c    Find minimum dxlmn of the cell sizes dxl(i) for the input mesh.
      c    Calculate the minimum allowable cell size dxmn for the output mesh.
      c    Get the endpoints of the interval (al,bpl) for the mesh.
      c    Set the initial value for the ratio raj of adjacent coarse mesh
      c    cell sizes to its maximum allowable value.  Set the initial value
      c    for the fine mesh cell size dxf to its minimum allowable value.
      c       Go to Step 3.
      c
 22. 30      continue
 23.         dxlmn = aminl(dxlmn,dxl(i))
 24.         i = i+1
 25.         if(i.le.imx) go to 30
 26.         dxmn = dxlmn/rmns
 27.         al = xl(0)
```

```
28.          bpl = xl(imx)
29.          raj = rajmx
30.          dxf = dxmn .
      c
      c       Step 3:
      c       Calculate bl and the number ncl of cells in a mesh for the
      c       subinterval (al,bl).
      c       If ncl =< imx, go to Step 4.
      c       Otherwise, repeat calculation of bl and ncl, increasing the value
      c       of dxf by multiplication with rdxf > 1.0 before each calculation,
      c       until ncl =< imx then go to Step 4, or until dxf exceeds its maximum
      c       allowable value then call HELP.
      c
31. 40       continue
32.          call segs(dxf,raj,clxf,crxf,nfs,xf,al,bpl,bl,ncl,imx,x(0),ial,ibl)
33.          if(ncl.le.imx) go to 60
34.          dxf = rdxf×dxf
35.          if(dxf.le.rdxmn×dxmn) go to 40
      c
      c       Call HELP.  DXF too big.
      c
36.          write(6,250)
37.          write(6,251)
38.          write(6,252) dxf,dxmn,ncl
39. 250      format(' DXF too big.')
40. 251      format(9x,'DXF',7x,'DXMN',4x,'NC1')
41. 252      format(1x,e11.5,e11.5,i7)
42.          call help(8hDXF 2big,8hSUB MESH)
      c
      c       Step 4:
      =       If ncl=imx, go to Step 7.
      c       Otherwise, calculate initial value of ratio rrj < 1.0 of
      c       successive trial values for raj.
      c       Initialize number nt of times through next loop.
      c       Go to Step 5.
      c
43. 60       continue
44.          if(ncl.eq.imx) go to 130
45.          frrj = frrjmx
46.          rrj = 1.0-frrj
47.          nt = 0
      c
      c       Step 5:
      c       Repeat calculation of bl and ncl, incrementing nt and decreasing
      c       raj by multiplication with rrj < 1.0 before each calculation, until
      c       ncl >= imx then go to Step 6, or until raj =< 1.0 then call HELP,
      c       or until nt >= ntmx then call HELP.
      c
48. 80       continue
49.          raj = amaxl(rrj×raj,1.0)
50.          if(nt.ge.ntmx) go to 110
51.          nt = nt+1
52.          call segs(dxf,raj,clxf,crxf,nfs,xf,al,bpl,bl,ncl,imx,x(0),ial,ibl)
53.          if(ncl.ge.imx) go to 90
54.          if(raj.le.1.0) go to 100
55.          go to 80
      c
      c
```

```
        c       Step 6:
        c       If nc1=imx, go to Step 7.
        c       Otherwise, change raj to its previous value before its last
        c     change.
        c       Calculate a larger value for rrj closer to 1.0, so that successive
        c     values of raj will be closer together.
        c       Go to Step 5.
        c
56. 90          continue
57.             if(nc1.eq.imx) go to 130
58.             raj = raj/rrj
59.             frrj = rfrrj*frrj
60.             rrj = 1.0-frrj
61.             go to 80
        c
        c       Call HELP.  RAJ too small.
        c
62. 100         continue
63.             write(6,300)
64.             write(6,301)
65.             write(6,302) raj,rrj,nt
66. 300         format(' RAJ too small.')
67. 301         format(9x,'RAJ',8x,'RRJ',3x,'NT')
68. 302         format(1x,e11.5,e11.5,i5)
69.             call help(8hRAJ 2sml,8hSUB MESH)
        c
        c       Call HELP.  NT too big.
        c
70. 110         continue
71.             write(6,310)
72.             write(6,301)
73.             write(6,302) raj,rrj,nt
74. 310         format(' NT too big.')
75.             call help(8hNT 2 big,8hSUB MESH)
        c
        c       Step 7:
        c       Calculate ib1, b1, and mesh x(i), (i=ia1,...,ib1) [where ia1=0],
        c     for subinterval (a1,b1).
        c       Compare ib1 with imx; if not equal, call HELP.
        c       Otherwise go to Step 8.
        c
76. 130         continue
77.             call xsegs(dxf,raj,clxf,crxf,nfs,xf,a1,bp1,b1,nc1,imx,x(0),ia1,
        *                  ib1)
78.             if(ib1.ne.imx) go to 140
        c
        c       Step 8:
        c       Calculate expansion factor rb, for expansion of all cell sizes in
        c     (a1,b1) to fill entire interval (a1,bp1) exactly.
        c       If expanded fine mesh cell size rb*dxf exceeds maximum allowable
        c     value, call HELP.
        c       Recalculate new expanded mesh x(i), (i=ia1,...,ib1), for entire
        c     interval (a1,bp1).
        c       Calculate cell sizes dx(i) and cell centers xc(i).
        c       Return.
        c
79.             rb = (bp1-a1)/(b1-a1)
80.             if(rb*dxf.gt.rdxmn*dxmn) go to 150
```

```
81.          do 135, i=1,imx
82.          x(i) = x(0)+(x(i)-x(0))*rb
83. 135      continue
84.          x(imx) = bp1
85.          do 136, i=1,imx
86.          dx(i) = x(i)-x(i-1)
87.          xc(i) = x(i-1)+dx(i)/2.0
88. 136      continue
89.          return
     c
     c       Call HELP.  IB1 not equal to IMX.
     c
90. 140      continue
91.          write(6,340)
92.          write(6,341)
93.          write(6,342) ib1,imx
94. 340      format(' IB1 not equal to IMX.')
95. 341      format(4x,'IB1',3x,'IMX')
96. 342      format(1x,i6,i6)
97.          call help(8hIB1neIMX,8hSUB MESH)
     c
     c       Call HELP.  RB*DXF too big.
     c
98. 150      continue
99.          write(6,350)
100.         write(6,351)
101.         write(6,352) rb,dxf,dxmn,rdxmn
102. 350     format(' RB*DXF too big.')
103. 351     format(10x,'RB',8x,'DXF',7x,'DXMN',6x,'RDXMN')
104. 352     format(1x,e11.5,e11.5,e11.5,e11.5)
105.         call help(8hRB*DXF2b,8hSUB MESH)
106.         end
```

```
1.          subroutine segs (dxf,raj,clxf,crxf,nfs,xf,a1,bp1,b1,nc1,
            *                 imx,x,ia1,ib1)
c
c    SUBROUTINE SEGS and ENTRY XSEGS                                    C.P.L.
c
c    PURPOSE of SEGS:
c
c       Given an input interval (a1,bp1), it outputs the right endpoint
c    b1 for a subinterval (a1,b1), and calculates the number nc1 of
c    cells in a mesh that partitions the subinterval into cells, giving
c    a fine mesh in places and a coarse mesh in the rest of the
c    subinterval.  The cells have the fine mesh cell size dxf in
c    the vicinity of each of the input points xf(i), (i=1,...,nfs).
c    Everywhere else, the cell sizes are as large (almost) as possible
c    consistent with the restriction that the ratio of adjacent cell
c    sizes (the larger divided by the smaller) is always between 1.0 and
c    raj.  The following requirements are satisfied:
c       (1) The right endpoint b1 of the subinterval is within a
c    distance dxf of the right endpoint bp1 of the input interval, ie.,
c    bp1-dxf < b1 =< bp1.  Also a1 =< b1 is required.
c       (2) The subinterval (a1,b1) is divided into segments, and each
c    segment is partitioned into cells.  There are two types of segments:
c    fine mesh segments and coarse mesh segments.  There are three kinds
c    of coarse mesh segments: the coarse mesh left segment, the coarse
c    mesh right segment, and coarse mesh middle segments.
c       (3) For a fine mesh segment, each cell has size dxf.
c       (4) For a coarse mesh left segment, the cell sizes are decreasing
c    from left to right by the ratio raj between adjacent cells.  The
c    last cell (at the right end) of a left segment has size rexp*dxf,
c    where rexp is required to satisfy 1 =< rexp =< raj.  A coarse mesh
c    left segment can appear only at the left end of the subinterval.
c       (5) For a coarse mesh right segment, the cell sizes are increasing
c    from left to right by the ratio raj between adjacent cells.  The
c    first cell (at the left end) of a right segment has size rexp*dxf,
c    where rexp is required to satisfy 1 =< rexp =< raj.  A coarse mesh
c    right segment can appear only at the right end of the subinterval.
c       (6) For a coarse mesh middle segment, the cell sizes are
c    increasing by the ratio raj between adjacent cells from the left
c    end to the center of the segment, and then are decreasing by the
c    ratio raj between adjacent cells from the center to the right end
c    of the segment.  The first and last cells (at the left and right
c    ends) of a middle segment have the same size rexp*dxf, where rexp
c    is required to satisfy 1 =< rexp =< raj.  A coarse mesh middle
c    segment can appear only in between two fine mesh segments.
c       (7) The value of rexp can be different for different coarse mesh
c    segments.
c       (8) Fine mesh segments are placed such that each point xf(i) is
c    covered by a fine mesh segment (unless b1 =< xf(i) =< bp1 ), and
c    there are at least 3.5 fine mesh cells between the point and the
c    nearest nontrivial coarse mesh segment on its left if any, and at
c    least 2.5 fine mesh cells between the point and the nearest
c    nontrivial coarse mesh segment on its right if any.
c       In the above, " a =< b " means "a less than or equal to b".
c
c    PURPOSE of XSEGS:
c
c       For a given input interval (a1,bp1), it does everything that
c    Subroutine SEGS does.  In addition, for the subinterval (a1,b1), it
```

```
c    outputs ibl and the mesh x(i), (i=ial,...,ibl), which partitions
c    the subinterval into cells by means of partitions located at
c    al=x(ial) < x(ial+1) < ... < x(ibl)=bl.  The cells (x(i-1),x(i)),
c    (i=ial+1,...,ibl), resulting from this partition satisfy the
c    requirements (2) through (8) listed above under PURPOSE of SEGS.
c       The integer ial is a constant set equal to 0 in Subroutine MESH.
c
c    INPUT ARGUMENTS for both SEGS and XSEGS:
c
c    dxf -- fine mesh cell size.
c    raj -- ratio of adjacent cell sizes in coarse mesh segments.
c    clxf -- minimum number of fine mesh cells to be placed between
c    each point xf(i) and the nearest coarse mesh segment on its left
c    if any.  This is a constant set equal to 3.5 in Subroutine MESH.
c    crxf -- minimum number of fine mesh cells to be placed between
c    each point xf(i) and the nearest coarse mesh segment on its right
c    if any.  This is a constant set equal to 2.5 in Subroutine MESH.
c    nfs -- number of points in the array xf.
c    xf(i), (i=1,...,nfs) -- points in the interval, which determine
c    neighborhoods in which all cells have the fine mesh cell size.
c    al -- left endpoint of the input interval (al,bpl).
c    bpl -- right endpoint of the input interval (al,bpl).
c
c    OUTPUT ARGUMENTS for both SEGS and XSEGS:
c
c    bl -- right endpoint of the subinterval (al,bl).
c    ncl -- number of cells in a mesh that partitions the
c    subinterval (al,bl).
c
c    INPUT ARGUMENTS for XSEGS only (Not used by SEGS):
c
c    imx -- maximum allowable subscript value for the array x.
c    ial -- first subscript value for the portion of the array x
c    containing the mesh for the subinterval (al,bl).  This is a
c    constant set equal to 0 in Subroutine MESH.
c
c    OUTPUT ARGUMENTS for XSEGS only (Not used by SEGS):
c
c    x(i), (i=ial,...,ibl) -- mesh which partitions the subinterval
c    (al,bl) into cells.
c    ibl -- last subscript value for the portion of the array x
c    containing the mesh for the subinterval (al,bl).
c
c    METHOD in SEGS:
c
c    Starting at the left endpoint of the input interval (al,bpl), it
c    places segments end to end inside the interval, one at a time, until
c    it reaches a point bl within a distance dxf of the right endpoint
c    bpl of the input interval.  The number nc of cells in each segment
c    is calculated when the segment is placed.  The sum of these numbers
c    over all segments gives the number ncl of cells in the subinterval
c    (al,bl).
c    The first segment to be placed is a coarse mesh left segment with
c    its left endpoint at al, and its right endpoint located at a
c    sufficient but limited distance to the left of the point xf(1).  If
c    there is insufficient room for a nontrivial left segment, then the
c    resulting segment has length 0, has no cells, and has both endpoints
c    at al.
```

```
c     Then it alternates between the placement of a fine mesh segment
c     and a coarse mesh middle segment until ncf fine mesh segments and
c     ncf-1 middle segments are placed.  Under usual circumstances the
c     i(th) fine mesh segment covers the point xf(i), however exceptions
c     can occur, and are mentioned in the next paragraph.
c        The i(th) fine mesh segment is placed with its left endpoint
c     coinciding with the right endpoint of the left segment if i=1 or
c     the right endpoint of the i-1(th) middle segment if i > 1, and its
c     right endpoint located at a sufficient but limited distance to the
c     right (under usual circumstances) of the point xf(i).  However, if
c     the point xf(i) lies within a distance dxf of the point bp1, it is
c     possible for the right endpoint of the fine mesh segment to be
c     located at the left of the point xf(i) within a distance dxf of the
c     point bp1.  There are also circumstances where the fine mesh
c     segment can be the trivial segment, with length 0 and having no
c     cells.  Under usual circumstances the left endpoint of the i(th)
c     fine mesh segment lies to the left of the point xf(i); however, if
c     the point xf(i) is already covered by a previous fine mesh segment
c     (one fine mesh segment might cover two or more of the points xf if
c     some of these points are closely spaced), then its left endpoint in
c     this case would lie to the right of the point xf(i).
c        The i(th) coarse mesh middle segment is placed with its left
c     endpoint coinciding with the right endpoint of the i(th) fine mesh
c     segment, and its right endpoint located at a sufficient but limited
c     distance to the left of the next point xf(i+1).  If there is
c     insufficient room for a nontrivial middle segment, then the
c     resulting segment has length 0 and has no cells.
c        Finally, a coarse mesh right segment is placed with its left
c     endpoint coinciding with the right endpoint of the last fine mesh
c     segment (fine mesh segment no. nfs), and its right endpoint b1
c     located as closely as possible to the point bp1 at its left.  If
c     there is insufficient room for a nontrivial right segment, then the
c     resulting segment has length 0, has no cells, and has both
c     endpoints at b1.
c        The subroutines LRSEG, FSEG, and MSEG are called to set up the
c     segments.  In each case, a confining interval (a,bp) for the
c     segment is used as the input interval to one of the subroutines.
c     The subroutine outputs the right endpoint b for a segment (a,b)
c     inside the input interval (a,bp), and calculates the number nc of
c     cells in the segment.
c        Before placing a coarse mesh left segment, it sets a=a1, and the
c     the point bp is placed at a distance of 3.5*dxf to the left of the
c     point xf(1).  Before placing the i(th) coarse mesh middle segment,
c     it sets a equal to the right endpoint b of the previous segment,
c     and the point bp is placed at a distance of at least 3.5*dxf to the
c     left of the point xf(i+1).  The term -(bp-b) solves a possible
c     iteration problem in Subroutine MESH.  Before placing the i(th)
c     fine mesh segment, it sets a equal to the right endpoint b of the
c     previous segment, and the point bp is placed at a distance of
c     3.5*dxf to the right of the point xf(i) if this location is at the
c     left of bp1, otherwise bp is placed at bp1.  Before placing the
c     coarse mesh right segment, it sets a equal to the right endpoint b
c     of the previous segment, and it places bp at bp1.  Finally it sets
c     b1 equal to the right endpoint b of the last segment placed (the
c     coarse mesh right segment).
c
c
c
```

```
         c     METHOD in XSEGS:
         c
         c        Given the input interval (a1,bp1), it does everything that
         c     Subroutine SEGS does, and in addition it calculates ib1 and the
         c     mesh x(i), (i=ia1,...,ib1), for the subinterval (a1,b1).
         c        Whenever it calls one of the subroutines LRSEG, FSEG, or MSEG to
         c     set up a segment (a,b) inside a confining interval (a,bp), it next
         c     provides a value for ia, and calls XLSEG, XFSEG, XMSEG, or XRSEG to
         c     calculate ib and the mesh x(i), (i=ia,...,ib), except x(ia), for
         c     that segment (a,b).  For a left segment, it sets ia=ia1.  For all
         c     other segments, it sets ia equal to the value of ib from the
         c     previous segment.  After the entire mesh is calculated, it sets ib1
         c     equal to the value of ib from the last segment to be placed (the
         c     right segment).
         c
         cXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
         c
  2.            dimension xf(nfs)
  3.            dimension x(0:imx)
         c
         c     Set ient=0 for Subroutine SEGS or ient=1 for Entry XSEGS.
         c
  4.            ient = 0
  5.            go to 10
  6.            entry xsegs(dxf,raj,clxf,crxf,nfs,xf,a1,bp1,b1,nc1,imx,x,ia1,ib1)
  7.            ient = 1
  8. 10         continue
         c
         c     Initialize fine cell count ncf and coarse cell count ncc.
         c
  9.            ncf = 0
 10.            ncc = 0
         c
         c     Step 1 (Place left segment):
         c     Initialize subscript i for xf array.
         c     Set a=a1 and find bp for confining interval (a,bp) for left
         c     segment.
         c        Calculate b and number nc of cells for left segment (a,b), and
         c     augment ncc.
         c        If in XSEGS, set ia=ia1, calculate ib and mesh x(j),
         c     (j=ia,...,ib), for left segment.
         c        Go to Step 3.
         c
 11.            i = 1
 12.            a = a1
 13.            bp = xf(i)-clxf*dxf
 14.            bp = amin1(bp,bp1)
 15.            call lrseg(dxf,raj,a,bp,b,nc,rexp)
 16.            ncc = ncc+nc
 17.            if(ient.eq.0) go to 30
 18.            ia = ia1
 19.            x(ia) = a
 20.            call xlseg(dxf,raj,b,nc,rexp,imx,x,ia,ib)
 21.            go to 30
         c
         c     Step 2 (Place middle segment):
         c     Increment subscript i for xf array.
         c     Set a equal to value of b from previous segment, and find bp for
```

```
        c   confining interval (a,bp) for middle segment.
        c     Calculate b and number nc of cells for middle segment (a,b), and
        c   -augment ncc.
        c     If in XSEGS, set ia equal to value of ib from previous segment,
        c   calculate new ib and mesh x(j), (j=ia,...,ib), except x(ia), for
        c   middle segment.
        c     Go to Step 3.
        c
22. 20      continue
23.         i = i+1
24.         a = b
25.         bp = xf(i)-clxf*dxf-(bp-b)
        c                                   The term -(bp-b) solves a possible
        c                                   iteration problem in Subroutine MESH.
26.         bp = amin1(bp,bp1)
27.         call mseg(dxf,raj,a,bp,b,nc,rexp)
28.         ncc = ncc+nc
29.         if(ient.eq.0) go to 30
30.         ia = ib
31.         call xmseg(dxf,raj,b,nc,rexp,imx,x,ia,ib)
        c
        c   Step 3 (Place fine segment):
        c     Set a equal to value of b from previous segment, and find bp for
        c   confining interval (a,bp) for fine segment.
        c     Calculate b and number nc of cells for fine segment (a,b), and
        c   augment ncf.
        c     If in XSEGS, set ia equal to value of ib from previous segment,
        c   calculate new ib and mesh x(j), (j=ia,...,ib), except x(ia), for
        c   fine segment.
        c     If more elements present in xf array, ie. i < nfs, go to Step 2.
        c     Otherwise go to Step 4.
        c
32. 30      continue
33.         a = b
34.         bp = xf(i)+(crxf+1.0)*dxf
35.         bp = amin1(bp,bp1)
36.         call fseg(dxf,a,bp,b,nc)
37.         ncf = ncf+nc
38.         if(ient.eq.0) go to 40
39.         ia = ib
40.         call xfseg(dxf,b,nc,imx,x,ia,ib)
41. 40      continue
42.         if(i.lt.nfs) go to 20
        c
        c   Step 4 (Place right segment):
        c     Set a equal to value of b from previous segment, and set bp=bp1
        c   for confining interval (a,bp) for right segment.
        c     Calculate b and number nc of cells for right segment (a,b), and
        c   augment ncc.
        c     Set b1=b, and calculate number nc1 of cells in subinterval
        c   (a1,b1).
        c     If in XSEGS, set ia equal to value of ib from previous segment,
        c   calculate new ib and mesh x(j), (j=ia,...,ib), except x(ia), for
        c   right segment, and set ib1=ib.
        c     Return.
        c
43.         a = b
44.         bp = bp1
```

```
45.          call lrseg(dxf,raj,a,bp,b,nc,rexp)
46.          ncc = ncc+nc
47.          bl = b
48.          ncl = ncf+ncc
49.          if(ient.eq.0) go to 50
50.          ia = ib
51.          call xrseg(dxf,raj,b,nc,rexp,imx,x,ia,ib)
52.          ibl = ib
53. 50       continue
54.          return
55.          end
```

```
1.           subroutine lrseg (dxf,raj,a,bp,b,nc,rexp)
     c
     c       SUBROUTINE LRSEG                                          C.P.L.
     c
     c       PURPOSE:
     c
     c          Given an input interval (a,bp), it outputs the right endpoint b
     c       for a coarse mesh left or right segment (a,b) inside the input
     c       interval, and calculates the number nc of cells in this segment.
     c       The following requirements are satisfied:
     c          (1) The right endpoint b of the segment is within a distance dxf
     c       of the right endpoint bp of the input interval, ie., bp-dxf < b
     c       =< bp.  Also a =< b is required. If bp =< a, then b=a and nc=0.
     c          (2) The last cell (at the right end) of a left segment or the
     c       first cell (at the left end) of a right segment has the size
     c       rexp×dxf, where the output argument rexp satisfies the inequality
     c       1 =< rexp =< raj.
     c          (3) For a left segment the cell sizes are decreasing from left
     c       to right by the ratio raj.  For a right segment the cell sizes
     c       are increasing from left to right by the ratio raj.
     c          In the above, " a =< b " means "a less than or equal to b".
     c
     c       INPUT ARGUMENTS:
     c
     c       dxf -- fine mesh cell size.
     c       raj -- ratio of adjacent cell sizes in coarse mesh segments.
     c       a -- left endpoint of the input interval (a,bp).
     c       bp -- right endpoint of the input interval (a,bp).
     c
     c       OUTPUT ARGUMENTS:
     c
     c       b -- right endpoint of the coarse mesh left or right
     c       segment (a,b).
     c       nc -- number of cells in the coarse mesh left or right
     c       segment (a,b).
     c       rexp -- ratio of the size of the last (at the right end) cell of
     c       the left segment to the size dxf of a fine mesh cell, or the ratio
     c       of the size of the first (at the left end) cell of the right segment
     c       to the size dxf of a fine mesh cell.
     c
     c       METHOD:
     c
     c          It makes a coarse mesh left or right segment that covers as much
     c       of the input interval as possible.
     c          It starts with a segment empty of cells, ie., the length ba=b-a
     c       of the segment (a,b) is zero.  Then it alternates between the two
     c       steps:
     c          (1) Increase the length ba of the segment by the addition of a
     c       cell of size dxf onto the right or left end of the left or right
     c       segment respectively.
     c          (2) Increase  the length ba of the segment by expanding the sizes
     c       of all cells in the segment by the factor raj.
     c          It continues the procedure of alternating between these two steps
     c       as long as the segment length ba remains less than the length
     c       bpa=bp-a of the input interval (a,bp).  If, when Step 1 is to be
     c       performed, but it would result in exceeding this length, the
     c       procedure is terminated, and it sets rexp=raj. If, when Step 2 is
     c       to be performed, but it would result in exceeding this length, the
```

```
      c   procedure is terminated, it sets rexp=bpa/ba, and an expansion of
      c   all cell sizes is made by the factor bpa/ba to make the segment
      c   fill the input interval exactly.
      c
      cXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
      c
  2.        bpa = bp-a
      c
      c       Initialize segment length ba and cell count nc.
      c
  3.        ba = 0.0
  4.        nc = 0
      c
      c       Step 1:
      c       Calculate value ba2 for tentative new segment length, if a cell
      c   is added to the segment.
      c       If tentative new segment does not fit inside the input interval,
      c   do not add cell, terminate procedure, and go to Step 3.
      c       If tentative new segment does fit inside the input interval, add
      c   a cell, and go to Step 2.
      c
  5. 10     ba2 = ba+dxf
  6.        if (ba2.gt.bpa) go to 20
  7.        ba = ba2
  8.        nc = nc+1
      c
      c       Step 2:
      c       Calculate value ba2 for tentative new segment length, if all
      c   cell sizes are expanded by the factor raj.
      c       If tentative new segment does not fit inside the input interval,
      c   do not expand cell sizes, terminate procedure, and go to Step 4.
      c       If tentative new segment does fit inside the input interval,
      c   expand cell sizes, and go to Step 1.
      c
  9.        ba2 = raj×ba
 10.        if (ba2.gt.bpa) go to 30
 11.        ba = ba2
 12.        go to 10
      c
      c       Step 3:
      c       Calculate rexp and b, and return.
      c
 13. 20     rexp = raj
 14.        b = a+ba
 15.        return
      c
      c       Step 4:
      c       Calculate rexp.
      c       Expand cell sizes by the factor bpa/ba to make the segment fill
      c   the input interval exactly.
      c       Calculate b, and return.
      c
 16. 30     rexp = bpa/ba
      c        ba = bpa    (not calculated)
 17.        b = bp
 18.        return
 19.        end
```

```
1.          subroutine xlseg (dxf,raj,b,nc,rexp,imx,x,ia,ib)
   c
   c        SUBROUTINE XLSEG                                              C.P.L.
   c
   c        PURPOSE:
   c
   c        For a coarse mesh left segment (a,b), it outputs all the mesh
   c     x(i), (i=ia,...,ib), except x(ia), which partitions the segment
   c     into cells by means of partitions located at a=x(ia) < x(ia+1)
   c     < ... < x(ib)=b.  The cells (x(i-1),x(i)), (i=ia+1,...,ib),
   c     resulting from this partition satisfy the requirements (2) and (3)
   c     in the listing for Subroutine LRSEG.
   c        The input arguments b, nc, rexp must be previously calculated
   c     by Subroutine LRSEG before entering Subroutine XLSEG.  Before
   c     calling Subroutine XLSEG, the Entry XSEGS to Subroutine SEGS gives
   c     a value to ia and sets x(ia)=a.
   c
   c        INPUT ARGUMENTS:
   c
   c        dxf -- fine mesh cell size.
   c        raj -- ratio of adjacent cell sizes in coarse mesh segments.
   c        b -- right endpoint of the coarse mesh left segment (a,b).
   c        nc -- number of cells in the segment.
   c        rexp -- ratio of the size of the last (at the right end) cell
   c     of the left segment to the size dxf of a fine mesh cell.
   c        imx -- maximum allowable subscript value for the array x.
   c        ia -- first subscript value for the portion of the array x
   c     containing the mesh for the left segment.
   c
   c        OUTPUT ARGUMENTS:
   c
   c        x(i), (i=ia,...,ib), [except x(ia), which is not part of the
   c     output] -- mesh which partitions the coarse mesh left segment
   c     into cells.
   c        ib -- last subscript value for the portion of the array x
   c     containing the mesh for the left segment.
   c
   c        UNUSED ARGUMENTS:
   c
   c        x(ia) -- the first element of the mesh for the left segment.
   c
   cXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
   c
 2.         dimension x(0:imx)
 3.         ib = ia+nc
 4.         if(nc.le.0) return
 5.         x(ib) = b
 6.         if(nc.le.1) return
 7.         dx = rexp*dxf
 8.         do 10, n=1,nc-1
 9.         x(ib-n) = x(ib-n+1)-dx
10.  10     dx = raj*dx
11.         return
12.         end
```

```
1.          subroutine xrseg (dxf,raj,b,nc,rexp,imx,x,ia,ib)
      c
      c     SUBROUTINE XRSEG                                          C.P.L.
      c
      c     PURPOSE:
      c
      c        For a coarse mesh right segment (a,b), it outputs all the mesh
      c     x(i), (i=ia,...,ib), except x(ia), which partitions the segment
      c     into cells by means of partitions located at a=x(ia) < x(ia+1)
      c     < ... < x(ib)=b.  The cells (x(i-1),x(i)), (i=ia+1,...,ib),
      c     resulting from this partition satisfy the requirements (2) and (3)
      c     in the listing for Subroutine LRSEG.
      c        The input arguments b, nc, rexp must be previously calculated by
      c     Subroutine LRSEG before entering Subroutine XRSEG.  Before calling
      c     Subroutine XRSEG, the Entry XSEGS to Subroutine SEGS sets ia equal
      c     to the value of ib from the previous segment that was formed.
      c     Since XSEGS sets a equal to the value of b from the previous
      c     segment that was formed, we will have x(ia)=a.
      c
      c     INPUT ARGUMENTS:
      c
      c        dxf -- fine mesh cell size.
      c        raj -- ratio of adjacent cell sizes in coarse mesh segments.
      c        b -- right endpoint of the coarse mesh right segment (a,b).
      c        nc -- number of cells in the segment.
      c        rexp -- ratio of the size of the first (at the left end) cell of
      c     the right segment to the size dxf of a fine mesh cell.
      c        imx -- maximum allowable subscript value for the array x.
      c        ia -- first subscript value for the portion of the array x
      c     containing the mesh for the right segment.
      c
      c     OUTPUT ARGUMENTS:
      c
      c        x(i), (i=ia,...,ib), [except x(ia), which is not part of the
      c     output] -- mesh which partitions the coarse mesh right segment
      c     into cells.
      c        ib -- last subscript value for the portion of the array x
      c     containing the mesh for the right segment.
      c
      c     UNUSED ARGUMENTS:
      c
      c        x(ia) -- the first element of the mesh for the right segment.
      c
      c********************************************************************
      c
2.          dimension x(0:imx)
3.          ib = ia+nc
4.          if(nc.le.0) return
5.          x(ib) = b
6.          if(nc.le.1) return
7.          dx = rexp*dxf
8.          do 10, n=1,nc-1
9.          x(ia+n) = x(ia+n-1)+dx
10. 10      dx = raj*dx
11.         return
12.         end
```

1.   subroutine mseg (dxf,raj,a,bp,b,nc,rexp)
```
c
c    SUBROUTINE MSEG                                              C.P.L.
c
c    PURPOSE:
c
c       Given an input interval (a,bp), it outputs the right endpoint b
c    for a coarse mesh middle segment (a,b) inside the input interval,
c    and calculates the number nc of cells in this segment.  The
c    following requirements are satisfied:
c       (1) The right endpoint b of the segment is within a distance dxf
c    of the right endpoint bp of the input interval, ie., bp-dxf < b
c    =< bp.  Also a =< b is required.  If bp =< a, then b=a and nc=0.
c       (2) The first and last cells (at the left and right ends) of a
c    middle segment have the same size rexp×dxf, where the output
c    argument rexp satisfies the inequality 1 =< rexp =< raj.
c       (3) For a middle segment, the cell sizes are increasing by the
c    ratio raj from the left end to the center of the segment, and then
c    are decreasing by the ratio raj from the center to the right end
c    of the segment.
c       In the above, " a =< b " means "a less than or equal to b".
c
c    INPUT ARGUMENTS:
c
c       dxf -- fine mesh cell size.
c       raj -- ratio of adjacent cell sizes in coarse mesh segments.
c       a -- left endpoint of the input interval (a,bp).
c       bp -- right endpoint of the input interval (a,bp).
c
c    OUTPUT ARGUMENTS:
c
c       b -- right endpoint of the coarse mesh middle segment (a,b).
c       nc -- number of cells in the coarse mesh middle segment (a,b).
c       rexp -- ratio of the size of the first or last (at the left or
c    right end) cell of a middle segment to the size dxf of a fine mesh
c    cell.
c
c    METHOD:
c
c       It makes a coarse mesh middle segment that covers as much of the
c    input interval as possible.
c       The segment (a,b) is regarded as being the union of two
c    subsegments, the left subsegment (a,c) and the right subsegment
c    (c,b).  It starts with both subsegments empty of cells, ie., the
c    lengths ca=c-a and bc=b-c of the two subsegments are zero.  Then
c    it rotates among the following three steps in order:
c       (1) Increase the length bc of the right subsegment by the
c    addition of a cell of size dxf onto the right end of the subsegment.
c       (2) Increase the length bc of the right subsegment by expanding
c    the sizes of all cells in the right subsegment by the factor raj.
c       (3) Compare the number of cells ncl and ncr in the left and right
c    subsegments.  If ncl < ncr, the sequence of cells in the combined
c    two subsegments are reversed in their order.  The subsegments
c    exchange positions, each keeping its original cells but in reversed
c    order.  Thus the values of ncl and ncr, which denote the number of
c    cells in the subsegment on the left and on the right respectively,
c    are exchanged.  Also, the values of ca and bc, which denote the
c    length of the subsegment on the left and on the right respectively,
```

```
      c   are exchanged.
      c      It continues the procedure of rotating among the three steps as
      c   long as the right subsegment length bc remains less than the length
      c   bpc=bp-c=bpa-ca of its confining interval (c,bp); or equivalently,
      c   as long as the length ba of the whole segment (a,b) remains less
      c   than the length bpa=bp-a of the input interval (a,bp).  If, when
      c   Step 1 is to be performed, but it would result in exceeding this
      c   length, the procedure is terminated, and it sets rexp=raj.  If, when
      c   Step 2 is to be performed, but it would result in exceeding this
      c   length, Step 2 is performed anyway, the procedure is then
      c   terminated, it sets rexp=raj*(bpa/ba), and a contraction of all
      c   cell sizes in both subsegments is made by the factor bpa/ba to make
      c   the whole segment fill the input interval exactly.
      c
      cxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
      c
 2.         bpa = bp-a
      c
      c      Initialize subsegment lengths ca and bc, and cell counts ncl and
      c   ncr.
      c
 3.         ca = 0.0
 4.         bc = 0.0
 5.         ncl = 0
 6.         ncr = 0
      c
 7.         bpc = bpa-ca
      c
      c      Step 1:
      c   Calculate value bc2 for tentative new right subsegment length, if
      c   a cell is added to the right subsegment.
      c      If tentative new right subsegment does not fit inside its
      c   confining interval, do not add cell, terminate procedure, and
      c   go to Step 4.
      c      If tentative new right subsegment does fit inside its confining
      c   interval, add a cell, and go to Step 2.
      c
 8. 10      bc2 = bc+dxf
 9.         if(bc2.gt.bpc) go to 20
10.         bc = bc2
11.         ncr = ncr+1
      c
      c      Step 2:
      c   Calculate value bc2 for tentative new right subsegment length, if
      c   all its cell sizes are expanded by the factor raj.
      c      If tentative new right subsegment does not fit inside its
      c   confining interval, do not expand cell sizes, terminate procedure,
      c   and go to Step 5.
      c      If tentative new right subsegment does fit inside its confining
      c   interval, expand all cell sizes in the right subsegment, and go to
      c   Step 3.
      c
12.         bc2 = raj*bc
13.         if(bc2.gt.bpc) go to 30
14.         bc = bc2
      c
      c      Step 3:
      c   If ncl < ncr, reverse the sequence of cells in the combined
```

```
      c    subsegments, thereby exchanging the left and right subsegments.
      c       Go to Step 1.
      c
15.          if(ncr.le.ncl) go to 10
16.          nexch = ncl
17.          ncl = ncr
18.          ncr = nexch
19.          sbexch = ca
20.          ca = bc
21.          bc = sbexch
22.          bpc = bpa-ca
23.          go to 10
      c
      c    Step 4:
      c    Calculate rexp, b, and number nc of cells in the whole segment
      c    (a,b), and return.
      c
24. 20       rexp = raj
25.          c = a+ca
26.          b = c+bc
27.          nc = ncl+ncr
28.          return
      c
      c    Step 5:
      c    Expand anyway, changing all cell sizes in the right subsegment
      c    by the factor raj.
      c    Calculate rexp.
      c    Contract all cell sizes in the whole segment (a,b) by the factor
      c    bpa/ba to make it fill the input interval exactly.
      c    Calculate b, and the number nc of cells in the whole segment,
      c    and return.
      c
29. 30       bc = bc2
30.          ba = bc+ca
31.          rexp = raj*bpa/ba
      c      ba = bpa    (not calculated)
32.          b = bp
33.          nc = ncl+ncr
34.          return
35.          end
```

```
1.        subroutine xmseg (dxf,raj,b,nc,rexp,imx,x,ia,ib)
      c
      c     SUBROUTINE XMSEG                                              C.P.L.
      c
      c     PURPOSE:
      c
      c        For a coarse mesh middle segment (a,b), it outputs all the mesh
      c     x(i), (i=ia,...,ib), except x(ia), which partitions the segment
      c     into cells by means of partitions located at a=x(ia) < x(ia+1)
      c     < ... < x(ib)=b.   The cells (x(i-1),x(i)), (i=ia+1,...,ib),
      c     resulting from this partition satisfy the requirements (2) and (3)
      c     in the listing for Subroutine MSEG.
      c        The input arguments b, nc, rexp must be previously calculated by
      c     Subroutine MSEG before entering Subroutine XMSEG.  Before calling
      c     Subroutine XMSEG, the Entry XSEGS to Subroutine SEGS sets ia equal
      c     to the value of ib from the previous segment that was formed.
      c     Since XSEGS sets a equal to the value of b from the previous
      c     segment that was formed, we will have x(ia)=a.
      c
      c     INPUT ARGUMENTS:
      c
      c        dxf -- fine mesh cell size.
      c        raj -- ratio of adjacent cell sizes in coarse mesh segments.
      c        b -- right endpoint of the coarse mesh middle segment (a,b).
      c        nc -- number of cells in the coarse mesh middle segment.
      c        rexp -- ratio of the size of the first or last (at the left or
      c     right end) cell of the middle segment to the size dxf of a fine
      c     mesh cell.
      c        imx -- maximum allowable subscript value for the array x.
      c        ia -- first subscript value for the portion of the array x
      c     containing the mesh for the middle segment.
      c
      c     OUTPUT ARGUMENTS:
      c
      c        x(i), (i=ia,...,ib), [except x(ia), which is not part of the
      c     output] -- mesh which partitions the coarse mesh middle segment
      c     into cells.
      c        ib -- last subscript value for the portion of the array x
      c     containing the mesh for the middle segment.
      c
      c     UNUSED ARGUMENTS:
      c
      c        x(ia) -- the first element of the mesh for the middle segment.
      cxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
2.        dimension x(0:imx)
3.        ib = ia+nc
4.        if(nc.le.0) return
5.        x(ib) = b
6.        if(nc.le.1) return
7.        nh = nc/2
      c                          Note: Integer division, giving a truncated result.
8.        dx = rexp*dxf
9.        do 10, n=1,nh
10.       x(ib-n) = x(ib-n+1)-dx
11.       x(ia+n) = x(ia+n-1)+dx
12. 10    dx = raj*dx
13.       return
14.       end
```

```
1.         subroutine fseg (dxf,a,bp,b,nc)
    c
    c      SUBROUTINE FSEG                                                C.P.L.
    c
    c      PURPOSE:
    c
    c      Given an input interval (a,bp), it outputs the right endpoint b
    c      for a fine mesh segment (a,b) inside the input interval, and
    c      calculates the number nc of cells in this segment.  The following
    c      requirements are satisfied:
    c      (1) The right endpoint b of the segment is within a distance dxf
    c      of the right endpoint bp of the input interval, ie., bp-dxf < b
    c      =< bp.  Also a =< b is required.  If bp =< a, then b=a and nc=0.
    c      (2) Each cell in the segment has size dxf.
    c      In the above, " a =< b " means "a less than or equal to b".
    c
    c      INPUT ARGUMENTS:
    c
    c      dxf -- fine mesh cell size.
    c      a -- left endpoint of the input interval (a,bp).
    c      bp -- right endpoint of the input interval (a,bp).
    c
    c      OUTPUT ARGUMENTS:
    c
    c      b -- right endpoint of the fine mesh segment (a,b).
    c      nc -- number of cells in the fine mesh segment (a,b).
    c
    c      METHOD:
    c
    c      It makes a fine mesh segment that covers as much of the input
    c      interval as possible.
    c      It starts with a segment empty of cells, ie., the length ba=b-a
    c      of the segment (a,b) is zero.  Then it attempts to repeat the
    c      following step: Increase the length ba of the segment by the
    c      addition of a cell of size dxf.  It continues the procedure of
    c      repeating this step as long as the segment length ba remains less
    c      than the length bpa=bp-a of the input interval (a,bp).  If when
    c      the step is to be started, but it would result in exceeding this
    c      length, the procedure is terminated.
    c
    cXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    c
2.         bpa = bp-a
    c
    c      Initialize segment length ba and cell count nc.
    c
3.         ba = 0.0
4.         nc = 0
    c
    c      Step 1:
    c      Calculate value ba2 for tentative new segment length, if a cell
    c      is added to the segment.
    c      If tentative new segment does not fit inside the input interval,
    c      do not add cell, terminate procedure, and go to Step 2.
    c      If tentative new segment does fit inside the input interval, add
    c      a cell, and go to start of Step 1.
    c
    c
```

```
 5. 10    ba2 = ba+dxf
 6.       if (ba2.gt.bpa) go to 20
 7.       ba = ba2
 8.       nc = nc+1
 9.       go to 10
    c
    c     Step 2:
    c     Calculate b, and return.
    c
10. 20    b = a+ba
11.       return
12.       end
```

```
 1.       subroutine xfseg (dxf,b,nc,imx,x,ia,ib)
    c
    c     SUBROUTINE XFSEG                                        C.P.L.
    c
    c     PURPOSE:
    c
    c       For a fine mesh segment (a,b), it outputs all the mesh
    c     x(i), (i=ia,....,ib), except x(ia), which partitions the segment
    c     into cells by means of partitions located at a=x(ia) < x(ia+1)
    c     < ... < x(ib)=b.  Each of the cells (x(i-1),x(i)), (i=ia+1,...,ib),
    c     resulting from this partition has size dxf.
    c       The input arguments b, nc must be previously calculated by
    c     Subroutine FSEG before entering Subroutine XFSEG.  Before calling
    c     Subroutine XFSEG, the Entry XSEGS to Subroutine SEGS sets ia equal
    c     to the value of ib from the previous segment that was formed.
    c     Since XSEGS sets a equal to the value of b from the previous
    c     segment that was formed, we will have x(ia)=a.
    c
    c     INPUT ARGUMENTS:
    c
    c     dxf -- fine mesh cell size.
    c     b -- right endpoint of the fine mesh segment (a,b).
    c     nc -- number of cells in the fine mesh segment (a,b).
    c     imx -- maximum allowable subscript value for the array x.
    c     ia -- first subscript value for the portion of the array x
    c     containing the mesh for the segment.
    c
    c     OUTPUT ARGUMENTS:
    c
    c     x(i), (i=ia,...,ib), [except x(ia), which is not part of the
    c     output] -- mesh which partitions the fine mesh segment into cells.
    c     ib -- last subscript value for the portion of the array x
    c     containing the mesh for the fine mesh segment.
    c
    c     UNUSED ARGUMENTS:
    c
    c     x(ia) -- the first element of the mesh for the fine mesh segment.
    c
    c*************************************************************************
    c
 2.       dimension x(0:imx)
 3.       ib = ia+nc
 4.       if(nc.le.0) return
 5.       x(ib) = b
 6.       if(nc.le.1) return
 7.       do 10, n=1,nc-1
 8.       x(ia+n) = x(ia+n-1)+dxf
 9. 10    continue
10.       return
11.       end
```

APPENDIX B

FORTRAN LISTING OF SUBROUTINE CMASS

1.   subroutine cmass(a,b,c,d,x1n,x2n)
```
c
c     SUBROUTINE CMASS                                    C. P. Luehr
c
c     PURPOSE:
c
c        This subroutine calculates the center of mass of the fluid portion
c     of a two dimensional partial island cell.  The output is to be used
c     by a finite difference fluid dynamics code which makes use of
c     partial island cells along the rigid boundaries of the fluid region.
c        With respect to a cartesian coordinate system (X,Y) in a plane,
c     the partial island cell is a square having its corners at the points
c     (0,0), (5,0), (0,5), and (5,5).  The square is divided into two
c     nontrivial parts by a straight line, which intersects the boundary
c     of the square at two points.  Each intersection point can either lie
c     on a corner or divide a side into two segments of integer lengths.
c     One part of the divided square is the fluid part, and the other is
c     the island part.  The fluid part is assumed to have a constant mass
c     density one.  The island part represents rigid solid material.
c        A description of the cell is given by a descriptor word, abcd,
c     where a, b, c, d are integers from 0 to 9 that describe the upper,
c     right, bottom, and left sides respectively of the square.  The
c     possible values of a and c correspond to the configurations shown
c     in Figure 1, where each unit segment of the form, o------o, bounds
c     on the fluid part of the cell, and each unit segment of the form,
c     o#####o, bounds on the island part of the cell.  Similarly, the
c     possible values of b and d correspond to the configurations in
c     Figure 2.  Examples of partial island cells are given in Figures 3,
c     4, 5, and 6.
c        The center of mass, x = (x1,x2), of the fluid part of the cell is
c     calculated and expressed in terms of normalized coordinates
c     (x1n,x2n), in terms of which the sides of the square have length
c     one.
c
c     INPUT ARGUMENTS:
c
c       a,b,c,d -- descriptors of the top, right, bottom, and left sides
c     respectively of the cell.  Each one is an integer from 0 to 9.
c
c     OUTPUT ARGUMENTS:
c
c       x1n,x2n -- the normalized cartesian coordinates of the center of
c     mass of the fluid portion of the cell in units for which the sides
c     of the cell have length one.
c
c     METHOD:
c
c        In every case, the fluid part of the cell is a 3, 4, or 5 sided
c     polygon, which can be obtained by removing a triangle from a
c     rectangle.  In some cases, the triangle may be a trivial triangle
c     with all its vertices lying on one line.  The area and center of
c     mass of the rectangle and of the triangle are found, and then used
c     for the calculation of the center of mass of the polygon.  The
c     procedure is as follows:
c        1.  First, it finds the corner q = (q1,q2) of the cell which is
c     at the greatest distance from the island part of the cell.  If the
c     two farthest corners from the island have the same distance from the
c     island, then q is chosen to be the one closest to the origin 0 =
```

```
c   (0,0) of the coordinate system.
c      If the length abs(d-5) of the fluid part of the left side is
c   greater than or equal to the length abs(b-5) of the fluid part of
c   right side, then q is one of the endpoints for the left side, ie.,
c   q1 = 0.  Otherwise, q is one of the endpoints for the right side,
c   ie., q1 = 5.  If the length abs(c-5) of the fluid part of the bottom
c   side is greater than or equal to the length abs(a-5) of the fluid
c   part of the top side, then q is one of the endpoints of the bottom
c   side, ie., q2 = 0.  Otherwise, q is one of the endpoints of the top
c   side, ie., q2 = 5.
c      2.  It finds the point p = (p1,p2) such that q and p are opposite
c   corners of the smallest rectangle containing all the fluid part of
c   the cell and having sides parallel to the sides of the cell.
c      If q1 = 0, then the fluid parts of the top and bottom sides, when
c   present, are on the left.  Then p1 equals the larger of the lengths
c   abs(a-5) and abs(c-5) of the top and bottom side fluid parts
c   respectively.  If q1 = 5, then the island parts of the top and
c   bottom sides, when present, are on the left.  Then p1 equals the
c   smaller of the lengths a and c of the top and bottom side island
c   parts respectively.  If q2 = 0, then the fluid parts of the right
c   and left sides, when present, are on the bottom.  Then p2 equals the
c   larger of the lengths abs(b-5) and abs(d-5) of the right and left
c   side fluid parts respectively.  If q2 = 5, then the island parts of
c   the right and left sides, when present, are on the bottom.  Then p2
c   equals the smaller of the lengths b and d of the right and left side
c   island parts respectively.
c      3.  It finds the points v = (v1,v2) and w = (w1,w2) on the
c   boundary of the cell separating the fluid and island parts of the
c   boundary, such that the sequence of points q, v, w defines a
c   counterclockwise sense of orientation.  The points v, p, w determine
c   a right triangle contained in the rectangle, such that the removal
c   of this triangle from the rectangle leaves the polygon which is
c   exactly the fluid part of the cell.
c      If q1 = 0, then v can only be on the left, bottom, or right side
c   of the cell.  Then v1 equals the length abs(c-5) of the bottom side
c   fluid part.  If q1 = 5, then v can only be on the right, top, or
c   left side.  Then v1 equals the length a of the top side island part.
c   If q2 = 0, then v can only be on the bottom, right, or top side.
c   Then v2 equals the length abs(b-5) of the right side fluid part.
c   If q2 = 5, then v can only be on the top, left, or bottom side.
c   Then v2 equals the length d of the left side island part.
c      If q1 = 0, then w can only be on the right, top, or left side of
c   the cell.  Then w1 equals the length abs(a-5) of the top side fluid
c   part.  If q1 = 5, then w can only be on the left, bottom, or right
c   side.  Then w1 equals the length c of the bottom side island part.
c   If q2 = 0, then w can only be on the top, left, or bottom side.
c   Then w2 equals the length abs(d-5) of the left side fluid part.  If
c   q2 = 5, then w can only be on the bottom, right, or top side.  Then
c   w2 equals the length b of the right side island part.
c      4.  It calculates the area at of the triangle, and the position
c   xt = (xt1,xt2) of the center of mass of the triangle.  The equation
c   for the center of mass in vector form is xt = (v+w+p)/3.
c      5.  It calculates the area ar of the rectangle, and the position
c   xr = (xr1,xr2) of the center of mass of the rectangle.  The equation
c   for the center of mass in vector form is xr = (q+p)/2.
c      6.  It calculates the center of mass x = (x1,x2) of the fluid
c   part, obtained by removing the triangle from the rectangle.  The
c   equation for the center of mass in vector form is
```

```
c    x = (ar*xr-at*xt)/(ar-at).   The result is expressed in terms of the
c    normalized coordinates (x1n,x2n), where x1n = x1/5, x2n = x2/5.
cxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
c
c    0  o-----o-----o-----o-----o-----o        o  o  o   o  o  o.  o  o  o  o
c                                                         |  |  :  #  #  #  #
c    1  .o#####o-----o-----o-----o-----o        |  |  |   |  #  #  #  #  #  #
c                                                         |  #  #  #  #  #  #
c    2  o#####o#####o-----o-----o-----o        o  o  o   o  o  o  o  o  o  o
c                                                         |  #  #  #  #  #  #
c    3  o#####o#####o#####o-----o-----o        |  |  |   #  #  #  #  #  #  #
c                                                         #  #  #  #  #  #  #
c    4  o#####o#####o#####o#####o-----o        o  o  o   o  o  o  o  o  o  o
c                                                         #  #  #  #  #  #  #
c    5  o#####o#####o#####o#####o#####o        #  #  |   #  #  #  #  #  #  #
c                                                         #  #  |  #  #  #  #
c    6  o-----o#####o#####o#####o#####o        o  o  o   o  o  o  o  o  o  o
c                                                         #  #  #  #  #  |  |
c    7  o-----o-----o#####o#####o#####o        |  |  #   #  #  #  #  |  |  |
c                                                         #  #  #  #  |  |  |
c    8  o-----o-----o-----o#####o#####o        o  o  o   o  o  o  o  o  o  o
c                                                         #  #  #  #  |  |  |
c    9  o-----o-----o-----o-----o#####o        |  #  #   #  #  #  |  |  |  |
c                                              o  o  o   o  o  o  o  o  o  o
c    /|\
c    /|\
c    :|:
c    a,c                           b,d ==>  0  1  2   3  4  5  6  7  8  9
c
c  Fig. 1  Top or Bottom Side.            Fig. 2  Right or Left Side.
c
cxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
c
c Y|                                    Y|
c  |                        w            |
c  qo-----o-----o-----o-----o#####o      o-----o-----o-----o-----o-----oq
c  |                      o#########      |                             |
c  |                    o###########      |                            |
c  |                  o#############       |                           |
c  o    FLUID       o###############o     vo       FLUID               o
c  |              o#################       ###                         |
c  |            o###################       #####o                      |
c  |          o#####################       ########                    |
c  o        o#######################o      o##########o                o
c  |      o#########################       ############                |
c  |    o###########################       ##############              |
c  vo##############################o  ###o ###############o            o
c  ################################## p #### o### ISLAND #########o
c  #############################(###### #### ##############################o
c  #######          #################### #### ############################
c  o####### ISLAND #################o    po#################################ow
c  ########################         ######################################
c  ########################         ######################################
c  ########################         ######################################
c Oo#####o#####o#####o#####o#####o-- Oo#####o#####o#####o#####o#####o--
c                              X                                     X
c
c  Fig. 3  Cell with abcd = 9552.        Fig. 4  Cell with abcd = 0154.
```
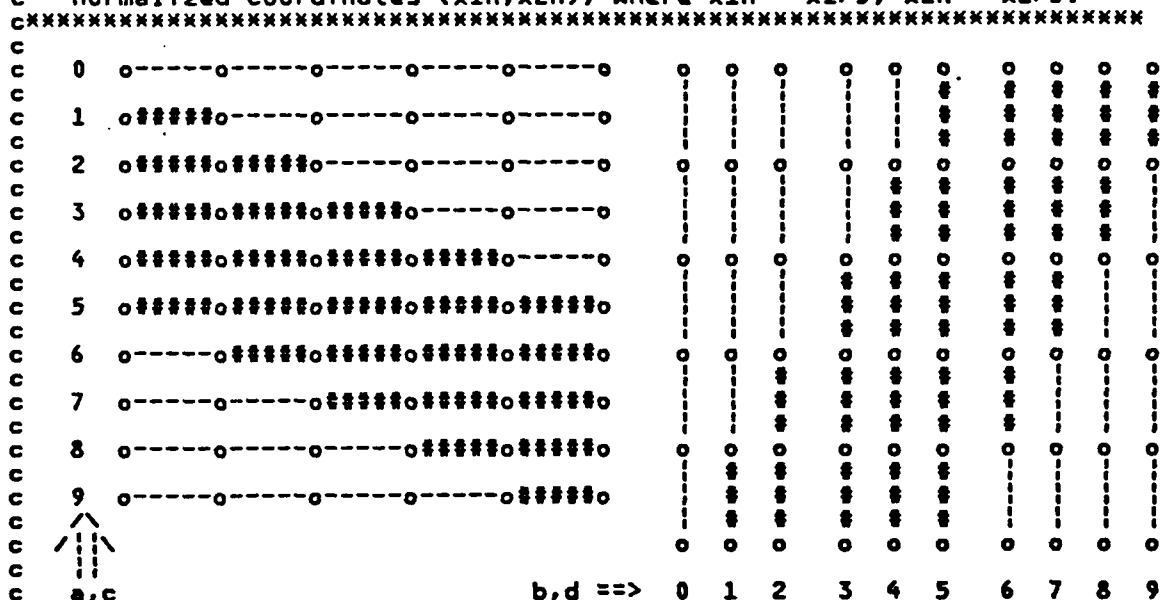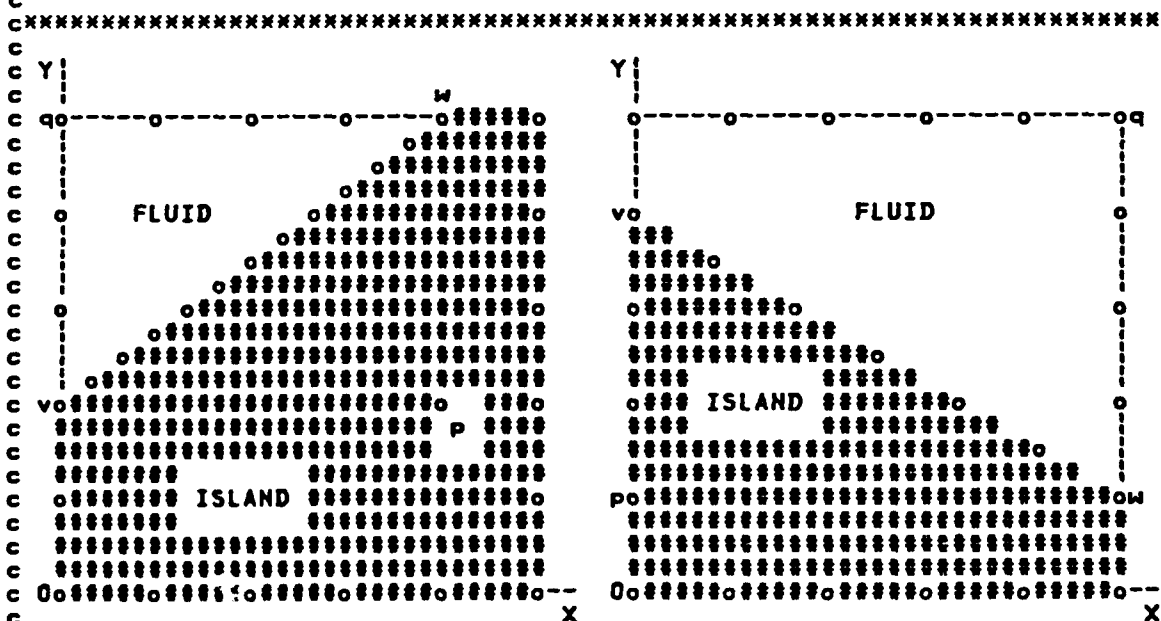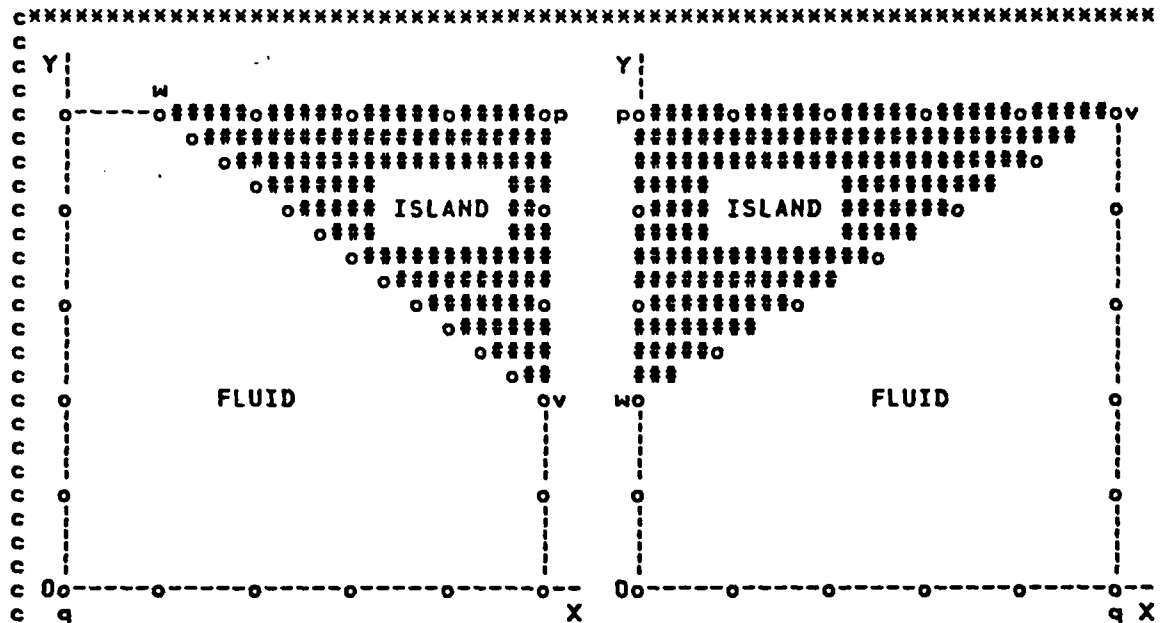
```
cxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
c
c  Y!                                        Y!
c   !                                         !
c   !      w                                  !
c   o-----o#####o#####o#####o#####op    po#####o#####o#####o#####o#####ov
c   !      o#####################         ###############################
c   !       o############################  ##############################o
c   !        o#########        ###          #####         ###########
c   o         o##### ISLAND ##o           o#### ISLAND #######o         o
c   !          o###         ###            #####         #####
c   !           o#################         #################o
c   !            o#############           #############
c   o             o########o              o#########o                   o
c   !              o#######o               #########
c   !               o#####                 #####o
c   !                o##                    ###
c   o         FLUID              ov    wo            FLUID               o
c   !                              !
c   !                              !
c   o                              o    o                                o
c   !                              !
c   !                              !
c  0o-----o-----o-----o-----o-----o--  0o-----o-----o-----o-----o-----o--
c  q                              X     q X
c
c  Fig. 5   Cell with abcd = 6700.      Fig. 6   Cell with abcd = 5007.
c
cxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
c
c
2.         integer a,b,c,d,q1,q2,p1,p2,v1,v2,w1,w2
c
c      Find the points q = (q1,q2) and p = (p1,p2).
c
3.         q1 = cvmgp(0,5,abs(d-5)-abs(b-5))
4.         q2 = cvmgp(0,5,abs(c-5)-abs(a-5))
5.         p1 = cvmgz(max(abs(a-5),abs(c-5)),min(a,c),q1)
6.         p2 = cvmgz(max(abs(b-5),abs(d-5)),min(b,d),q2)
c
c      Find the points v = (v1,v2) and w = (w1,w2).
c
7.         v1 = cvmgz(abs(c-5),a,q1)
8.         v2 = cvmgz(abs(b-5),d,q2)
9.         w1 = cvmgz(abs(a-5),c,q1)
10.        w2 = cvmgz(abs(d-5),b,q2)
c
c      Find the area at and the center of mass xt = (xt1,xt2) of the
c      triangle.
c
11.        at = 0.5*real(abs(v1-w1)*abs(v2-w2))
12.        xt1 = real(v1+w1+p1)*0.33333333
13.        xt2 = real(v2+w2+p2)*0.33333333
c
c      Find the area ar and the center of mass xr = (xr1,xr2) of the
c      rectangle.
c
```

```
14.          ar = real(abs(p1-q1)*abs(p2-q2))
15.          xr1 = real(q1+p1)*0.5
16.          xr2 = real(q2+p2)*0.5
      c
      c      Calculate the center of mass x = (x1,x2) of the fluid part of the
      c      cell, and express the result in normalized coordinates (x1n,x2n),
      c      where x1n = x1/5 and x2n = x2/5.
      c
17.          ra = 0.2/(ar-at)
18.          x1n = (ar*xr1-at*xt1)*ra
19.          x2n = (ar*xr2-at*xt2)*ra
20.          return
21.          end
```

1983-1984 AFWL-SCEEE WEAPONS LABORATORY SCHOLAR PROGRAM

Sponsored by the

AIR FORCE WEAPONS LABORATORY

Conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

FINAL REPORT

INTRACAVITY ABSORPTION AND GAIN DETECTION

Prepared by: Dr. Rodney L. Williamson

Research Location: Air Force Weapons Laboratory,
                         Advanced Radiation Division,
                         Short Wavelength Laser Branch

AFWL Research Cantact: Dr. Steven J. Davis

Date: September 16, 1984
Contract Number: F49620-82-C-0035

INTRACAVITY ABSORPTION AND GAIN DETECTION

I. INTRODUCTION

When developing new laser systems, it is of fundamental importance to have at one's disposal a sensitive method of gain detection. If one is planning to investigate a number of potential laser systems, it is convenient if this gain detection method has the versatility to be equally applied in all cases. The intracavity gain detection method is such a technique. It has been shown to be very sensitive, capable of measuring gains (or absorptions) as small as 10^{-4}, and can be equally well applied to both CW and pulsed systems. Intracavity absorption and gain measurement techniques were pioneered in the early 1970's.[1-6] All of these techniques involve placing the gain or loss medium directly inside the cavity of an operating CW or pulsed dye laser.

The enhancement of sensitivity relative to extracavity techniques is due primarily to two factors.[7] First, a photon traveling inside the cavity will travel back and forth between the resonator mirrors an average of $1/T_{out}$ times, where T_{out} is the transmittance of the output coupler. The probability that the photon will be absorbed (or amplified) is, therefore, also increased by this factor relative to a conventional extracavity single pass method. Secondly, at threshold, the unsaturated gain in the probe laser medium exactly offsets the cavity losses. Thus, any gain or loss introduced by placing the test medium inside the resonator upsets this delicate balance either enhancing or attenuating the output beam. Under these conditions, the probe laser becomes extremely sensitive to small losses or gains introduced into the cavity.

An expression may be derived[7] describing the relative change in the probe laser output power due to the introduction of a small loss or gain, g. We have

$$\Delta I/I = G_o I_s [L_c(1+L_c/g)]^{-1}[I_s(G_o/L_c-1)+I_b]^{-1} \qquad (1)$$

where G_o is the unsaturated gain, I_s is the satur-
ation intensity, L_c represents the cavity losses, and I_b is
the background intensity due to spontaneous emission. It is
easily seen that in the threshold limit, this expression
reduces to

$$\lim_{G_o \to L_c} (\Delta I/I) = I_s g/(L_c+g) I_b \qquad (2)$$

Since, in general, $I_s \gg I_b$, it is seen that near threshold
the laser becomes very sensitive to g. Note that this
equation is often derived under the assumption that I_b may
be ignored.[7] In this limit equation (2) goes to infinity.

The requirement of versatility is achieved by using a
dye laser as the probe laser which allows tuning to dif-
ferent wavelengths where gain or absorptions are to be
detected. Thus, one set-up may be used to detect small gains
or losses over the entire spectral region covered by the
particular dye laser in use. Equally important in gain
detection experiments is the observation that the wavelength
of the probe laser output will "pull" or "lock onto" the
fluorescence wavelength of the inverted test medium.[1-3]
This is illustrated in Figure 1, which depicts the
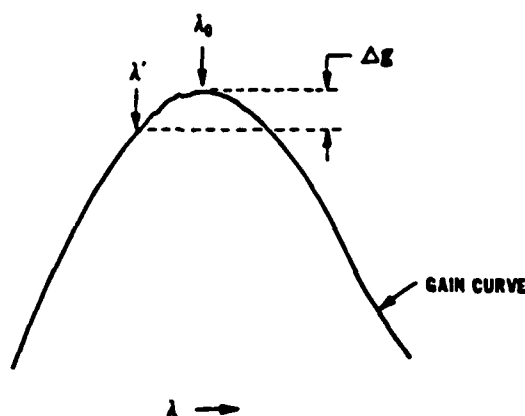
FIGURE 1 - Pulling the probe laser output.

transmittance curve of the intracavity tuning element, in this case a Lyot filter. The probe laser oscillates at λ_0, the wavelength of maximum transmittance of the tuning element. (The point of maximum transmittance may be tuned by rotating the Lyot filter about an axis normal to the filter plane.) If gain is introduced at some other wavelength, say λ', then the laser will oscillate at this new wavelength provided the gain is sufficiently large to overcome the tuning element loss. This circumvents the troublesome necessity of having to "perfectly" match the probe laser wavelength to the gain medium flourescence. How far the probe laser can be "pulled" is a function of the magnitude of the gain introduced by the test medium.

II. OBJECTIVE

At a research facility such as AFWL where there is an ongoing interest in the development of new laser systems, it is important, and extremely convenient, to have ready access to a sensitive and versatile gain detection technique. The technique which most successfully fulfills these two criteria is the intracavity gain detection technique.

The principle objective of this research effort was to develop and test a sensitive gain measurement technique to be used on candidate visible laser systems which were potentially capable of being chemically pumped. The test system chosen was iodine monofluoride, a system which has not yet been lased upon chemical pumping but which shows considerable promise.[8] We hoped to illustrate that the gain diagnostic was sufficiently sensitive to detect subthreshold gain, and that it could be used to optimize conditions to obtain threshold gain values. The primary research objective was broken up into a number of subobjectives. First, preliminary experiments were to be done to determine the sensitivity of the intracavity technique relative to cavity

lifetime measurements. The cavity lifetime is a measure of the average time a photon requires to "leak" out of the optical cavity once it has been injected. Thus, if there is a gain/loss medium in the cavity, the lifetime will be lengthened/shortened relative to the empty cavity. Small changes in cavity lifetime can be detected through amplitude modulation techniques.[9,10] These will be discussed below. Secondly, we wanted to use whichever technique was found to be most desirable to detect optically pumped gain on some known systems. We proposed to characterize the technique not only on CW systems, as had been done earlier by other workers, but also on pulsed systems. Thirdly, we wanted to use the gain detection technique as an optimization tool on a chemically pumped system.

III. INTRACAVITY MOLECULAR IODINE ABSORPTION

To test the sensitivity of the intracavity detection technique the experimental set-up shown in Figure 2 was
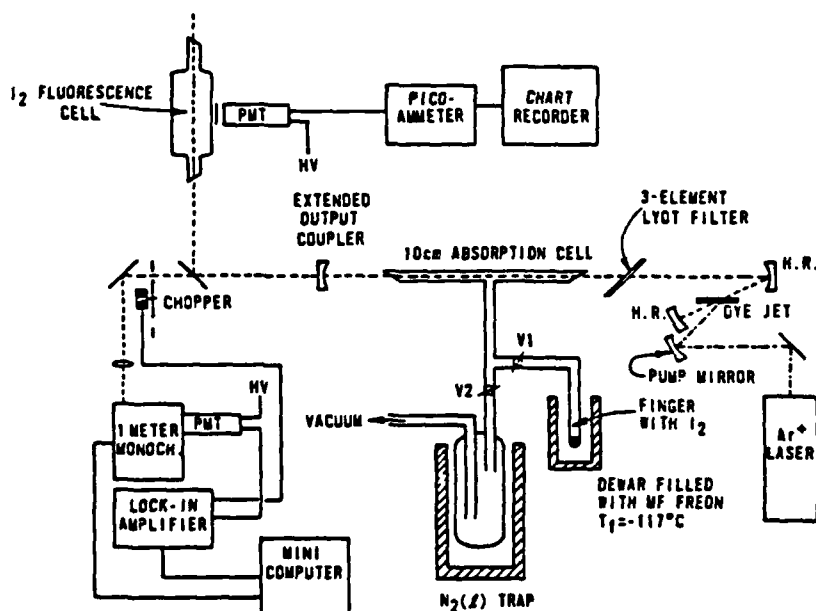
FIGURE 2 - Intracavity absorption apparatus.

assembled. The test system chosen was molecular iodine. The dye used in the tunable CR-590 CW dye laser (Coherent) was rhodamine 590 and the dye laser was pumped with the 514.5 nm output (multimode) of an Ar^+ laser (Spectra Physics Mo. 164). The output coupler of the dye laser was extended 20 cm to incorporate the I_2 absorption cell. The I_2 concentration in the cell could be controlled by adjusting the temperature of a cold bath (freon) in which a side-arm containing iodine crystals was immersed. The I_2 vapor pressure could be adjusted from approximately 300 mtorr down to less than 10^{-9} torr.

The I_2 absorption lines which figure prominently in the spectral region of the dye at room temperature (the temperature of the gas in the absorption cell) are shown in Table 1, along with their relative intensity factors. The relative intensity factor for a particular transition is given by the product of the Franck-Condon[11] and Boltzmann factors, multiplied by 100. The calculated values for the I_2 bandhead positions were obtained from a computer program written by S. J. Davis. The experimental values were obtained from an excitation spectrum taken with the apparatus shown in Figure 2, with the intracavity cell evacuated.

A typical intracavity I_2 absorption spectrum was taken as follows. The gas in the intracavity cell was allowed to come to thermal equilibrium with the liquid nitrogen cold trap. The dye laser was tuned to an I_2 absorption line (by noting the fluorescence maximum) and the pump power adjusted so that the laser was operating just far enough above threshold so that the output was stable. The steady-state fluorescence level was then recorded and used as the baseline. To maximize the sensitivity with which small baseline changes could be detected, the offset option on the chart recorder amplifier was used and the gain optimized.

v'-v"	WL(obs)	WL(cal)	I.F.		v'-v"	WL(obs)	WL(cal)	I.F.
14-1	584.4	584.5	0.47		9-2	611.1	611.2	0.12
16-2	584.8	584.9	0.31		7-1	611.5	611.5	0.022
13-1	588.0	588.1	0.37		12-4	614.6	614.7	0.021
15-2	588.2	588.4	0.33		10-3	614.8	614.9	0.12
12-1	591.7	591.7	0.27		8-2	615.4	615.3	0.076
14-2	591.7	591.9	0.33		11-4	618.7	618.8	0.034
11-1	595.3	595.5	0.19		9-3	619.0	619.1	0.10
13-2	595.3	595.5	0.31		7-2	619.5	619.6	0.044
12-2	599.3	599.2	0.27		10-4	622.8	622.9	0.044
10-1	599.3	599.3	0.13		8-3	623.3	623.4	0.077
13-3	602.9	603.1	0.11		6-2	624.0	624.0	0.023
11-2	603.0	603.1	0.22		9-4	627.0	627.2	0.048
9-1	603.1	603.3	0.077		7-3	627.7	627.8	0.052
12-3	606.7	606.8	0.13		8-4	631.4	631.6	0.046
10-2	607.0	607.0	0.17		6-3	632.1	632.3	0.031
8-1	607.1	607.3	0.043		7-4	635.9	636.2	0.038
11-3	610.7	610.8	0.13		5-3	636.8	637.0	0.017

TABLE 1 - Iodine absorption lines and the associated intensity factors in the WL region of the R590 gain curve.

Valve V2 was then closed, isolating the absorption cell from the liquid nitrogen trap, and V1 was opened and time allowed for the system to come to equilibrium. Intracavity absorptions were detected by monitoring corresponding intensity changes in the flourescence. Typical data obtained in such experiments is shown in Figure 3. I_2 vapor pressures of approximately 7×10^{-6} torr could be detected in this manner at 599.2 nm. Assuming simultaneous absorption by the overlapping 10'-1" and 12'-2" bands of the B-X transition, this corresponds to detection of less than 10^{10} I_2 molecules.
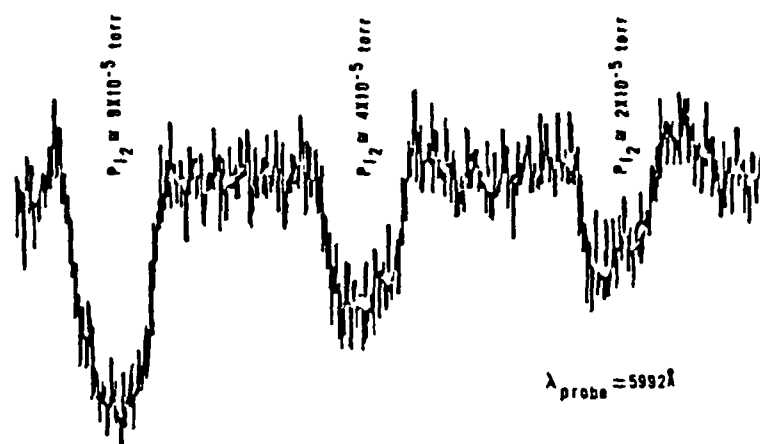
FIGURE 3 - Intracavity molecular iodine absorptions.

Intracavity molecular iodine absorptions were also detected directly by observing changes in the dye laser output line shape as a function of the I_2 vapor pressure in the absorption cell. The laser line shape was detected with a 1 meter scanning monochromator (GCA MacPherson Mo. 2051) in second order. Typical data are shown in Figure 4 for two different bath temperatures. Note the large decrease in relative intensity when the I_2 vapor is let into the absorption cell.
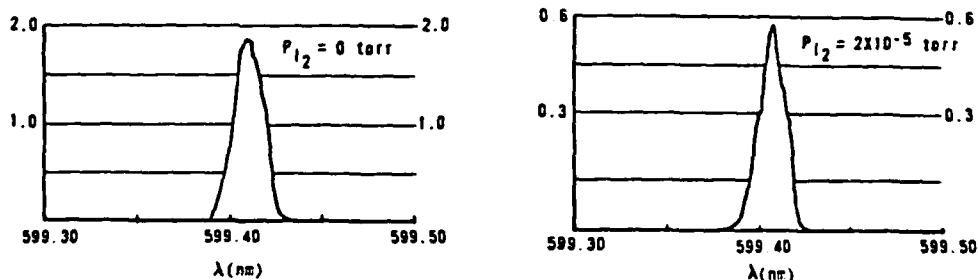
FIGURE 4 - The effect of intracavity absorption on the probe laser line shape.

To compare the above results with those obtained from an extracavity method[9,10] the apparatus shown in Figure 5 was set up to measure changes in the cavity lifetime due to molecular iodine absorptions. The phase of an amplitude
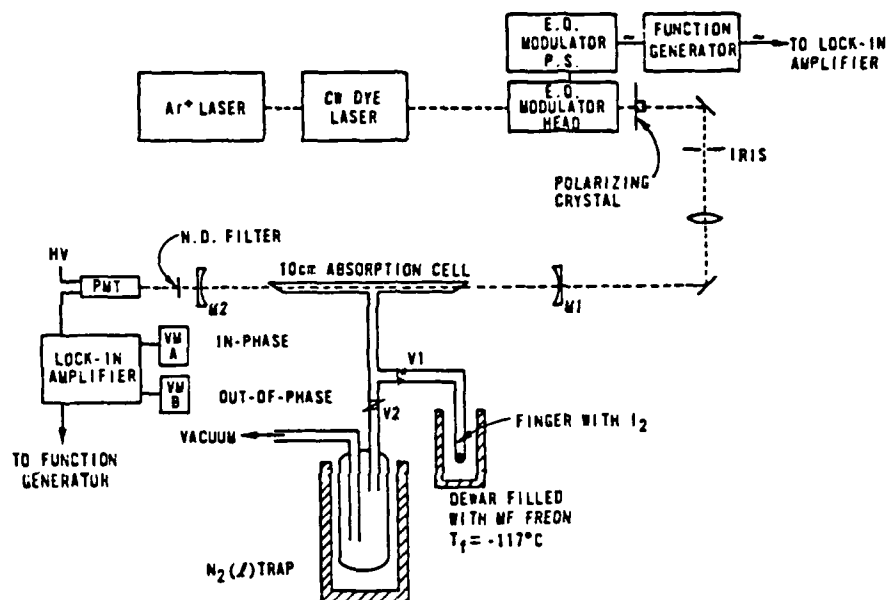
FIGURE 5 - Phase modulation spectroscopy apparatus.

modulated beam will be shifted after the beam has passed through an optical cavity. (This is the optical analog to the phase shift introduced into an ac voltage upon passing through an RC circuit.) The phase shift is given by

$$\phi = \tan^{-1} Tf \tag{3}$$

where T is the cavity lifetime and f is the modulation frequency. The change in the phase shift, ϕ', induced by introducing a small loss or gain, g, into the cavity is given by

$$\tan\phi' = -cg \, \tan\phi / [cg + L_t (1 + \tan^2 \phi)] \tag{4}$$

where c is the speed of light and L_t is the intensity loss per unit time of the cavity. With the modulation frequency set at 200 kHz changes in the phase shift could be detected for iodine vapor pressures of 8×10^{-4} torr. Thus, with the

set-up shown, the cavity lifetime method was nearly two orders of magnitude less sensitive than the intracavity technique. (This gap can probably be narrowed somewhat under more ideal circumstances. The Coherent Model 317 electro-optic modulator did not yield sinusoidal waveforms at these frequencies, the waveform was more that of a sawtooth. Presumably, this would introduce high frequency Fourier components into the beam which would change the nature of the phase shifts. Additional problems were encountered due to a large decrease in modulation amplitude with increasing frequency.)

IV. PULSED GAIN DETECTION

The apparatus shown in Figure 6 was assembled to detect pulsed gain on various visible transitions of I_2 and IF. There are certain aspects of the apparatus which are crucial to the success of the measurement and which, therefore, need to be discussed. First, the stock 3-element Lyot filter supplied with the CR-590 dye laser was modified so as to consist of only the single thickest quartz plate. The

FIGURE 6 - Pulsed intracavity gain detection apparatus.

desired effect was to decrease the finesse of the filter. This enables a given amount of gain (created in the cell by the optical pump pulse) to pull the wavelength of the probe laser output further because it has a smaller filter loss to overcome (see the discussion in the INTRODUCTION). An undesirable side effect is that this allows the filter to pass more than one wavelength, each wavelength corresponding to a different order of the filter. It is thus possible to tune the single element Lyot filter such that two, or sometimes even three, wavelengths oscillate simultaneously, each separated by about 0.3 nm. (With the 3-element filter the finesse is sufficiently high that all but one order is suppressed.) This also causes the laser to tune discretly, from order to order, rather than continuously.

Another important element of the apparatus is the long pass filter mounted at Brewster's angle to the optical axis on the tuning element housing. This filter is necessary because the Lyot filter is not sufficiently selective to block the pump laser pulse. The pump power used to obtain optically pumped gain for both I_2 and IF is sufficient to vaporize the dye jet if the long pass filter is not inserted, thus severely interrupting the probe laser output. This interruption occurs over a time period of hundreds of microseconds, and completely masks the gain signal. It is worth noting here that insertion of the filter is not straightforward and a fair amount of effort is usually required to achieve an orientation which introduces a sufficiently small loss to allow the laser to come to threshold. This orientational dependence of the loss introduced by the long pass filter is not understood. Also, the pump laser pulse bleaches the filter and its subsequent recovery introduces a time dependent loss into the cavity. However, this effect is not sufficiently large to interfere with detection of the gain signal. It might be supposed that

the need to insert the long pass filter can be circumvented by pumping off axis with the pump laser. This was attempted with little success. The technique is very sensitive to the overlap of the probe and pump beams, and the magnitude of the gain signal decreases too rapidly as the pump beam is moved off axis to make such a solution viable.

The apparatus depicted in Figure 6 was used to detect optically pumped gain on various flourescent transitions of IF as follows. The probe laser was brought to threshold and used to line up the pump laser beam on axis. (It is very important to obtain maximum overlap of the two beams.) Iodine and fluorine were then admitted into the intracavity cell and the flows adjusted to maximize the IF chemilumi-nescence at the desired total pressure. The probe laser was tuned to the desired wavelength region with the aid of the OMA (labeled detection system 3 in the figure). Detection system 1 was then tuned to the bandhead of the fluorescence transition that was to be probed for gain. The test medium was then optically excited with the pump laser at a wave-length of about 496.5 nm, which corresponds to pumping the $B(3')-X(0'')$ absorption band of IF. (The 2'-0" and 4'-0" transitions were also frequently used.) The pump laser was fine tuned to maximize the bright yellow side fluorescence detected with detection system 2. Pulse powers ranged from 20 to 50 mjoules/pulse. The pulse rate was typically 0.3 Hz. At this point gain pulses were usually obtained on both detection systems 1 and 4. The spectrally resolved gain was detected on system 1 and the unresolved gain on system 4. Scanning detection system 1 revealed that the gain pulse did not come at any one particular wavelength but was spread over a broad region of the band. The maximum width of the region to which the dye laser could be pulled was typically about 1.5-2 nm. The monochromator was usually scanned in this region to find the wavelength where the gain was at its

maximum. This region became increasingly narrow as the probe
laser was moved further from the gain transition wavelength.
The unresolved gain could be maximized by tuning the probe
laser. If gain pulses were not immediately obtained
adjustment of the gas flows or pump wavelength was usually
required.

Figure 7 shows examples of data obtained in a pulsed
gain detection experiment. The upper traces show the gain
pulse and side fluorescence associated with the thermalized
B(0')-X(5") transition at 625.1 nm at a cell pressure of 32
torr. For this particular spectrum the probe laser was set
at 621.5 nm and detection system 1 was set to 625.4 nm (1 nm
bandpass). The lower traces show the gain pulse and side
fluorescence associated with the direct B(3')-X(7") transi-
tion at 622.9 nm. The total cell pressure was 2.5 torr. The
probe laser was set at 621.8 nm and the gain was detected at
623.5 nm. It was found that the direct gain disappeared at a
total cell pressure of about 4 torr and that thermalized[13]
gain onset at about 5 torr. Attempts to detect both direct
and thermalized gain at the same cell pressure were not
successful. The dye laser pump beam was blocked and it was
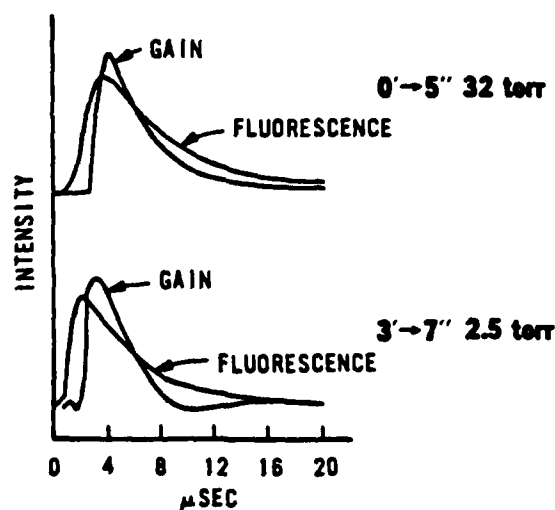
FIGURE 7 - Thermalized and direct IF gain pulses.

confirmed that the optically pumped IF could not be lased without the assistance of the dye laser for either transition. Thus, for the dye laser cavity the gain detected was subthreshold. Note that lasing has previously been achieved on both of these transitions.[12,13]

Note that the spectra shown in Figure 7 were taken with a "slow" plug-in amplifier (HP 7A22 - 1 MHz bandpass, 350 ns rise time) in the HP 7844 oscilloscope, and do not show the accurate time evolution of the gain or fluorescence pulses. The slow plug-in was chosen because of its sensitivity; its use made the gain pulses very easy to detect. To follow the time evolution of the pulses an HP 7A19 plug-in amplifier was used (400 MHz bandpass, 0.9 ns rise time). To gain sufficient intensity the signal had to be amplified before it was fed into the oscilloscope, and a HP 461A amplifier was used for this preamplification. Figure 8 shows an example of a thermalized B(0')-X(5") gain pulse detected with this "fast" system. The cell pressure was 21 torr. The probe laser was set at 620.5 nm and the gain was detected at 625.2 nm (1 nm bandpass).

IF B(0')-X(5") GAIN PULSE

I$_2$ B-X GAIN PULSE

$P_{cell} = 21$ torr
$\lambda_{probe} = 6205$Å
$\lambda_{mon} = 6252$Å
PUMPED AT ~4965Å

~ 20 mtorr I$_2$
$\lambda_{probe} = 5982$Å
$\lambda_{mon} = 5946$Å
PUMPED AT ~5129Å

0 2 4 6 8 10
μsec

0   1   2
μsec
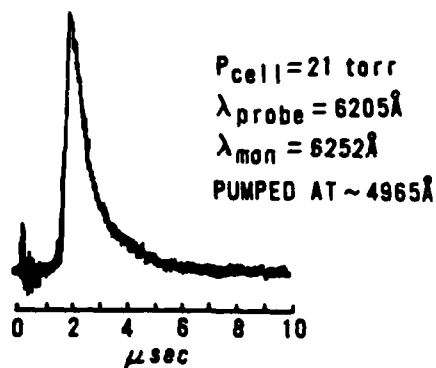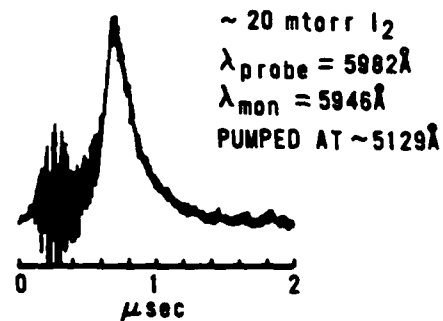
FIGURE 8 - Example of optically pumped IF and I$_2$ gain pulses

The pulsed IF side fluorescence was briefly studied using D.S.2 of the gain detection apparatus (Figure 6) with the "fast" electronics option described above. Figure 9 shows a plot of the intensity rise time of the fluorescence pulse versus v', the vibrational level from which the fluorescence originates. The IF was excited into the B(3') state at 496.5 nm. The B-X transitions examined were the 3'-11", 2'-9", 1'-7", and 0'-4". The total cell pressure was held constant at 33 torr. The increase in the time required for the onset of fluorescence with decreasing v' reflects the longer time required for the vibrational levels further away from the initially populated level to achieve population through thermalization processes. Fluorescence emission from vibrational levels with v'>3 was not observed.
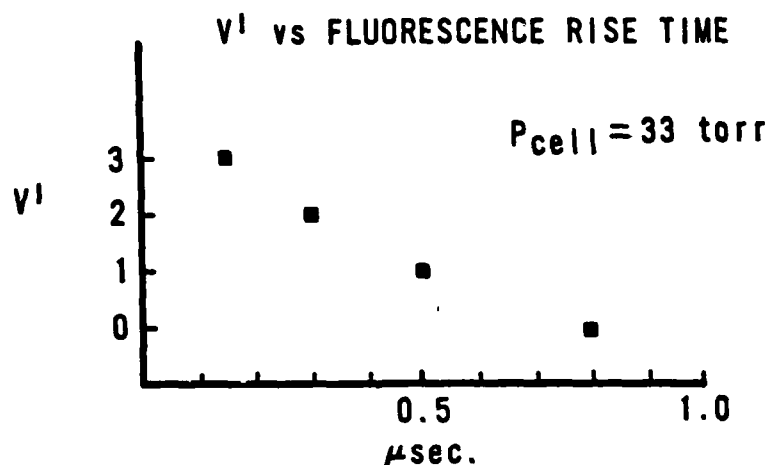
FIGURE 9 - IF B-X fluorescence rise time as a function of v'

Figure 10 shows a plot of the intensity rise time of the 2'-5" fluorescence as a function of the total cell pressure. Here again we see the expected trend. As the pressure is increased the rise time becomes shorter due to the fact that the collision frequency, and thus the rate of thermalization down to the 2' level, is increased.
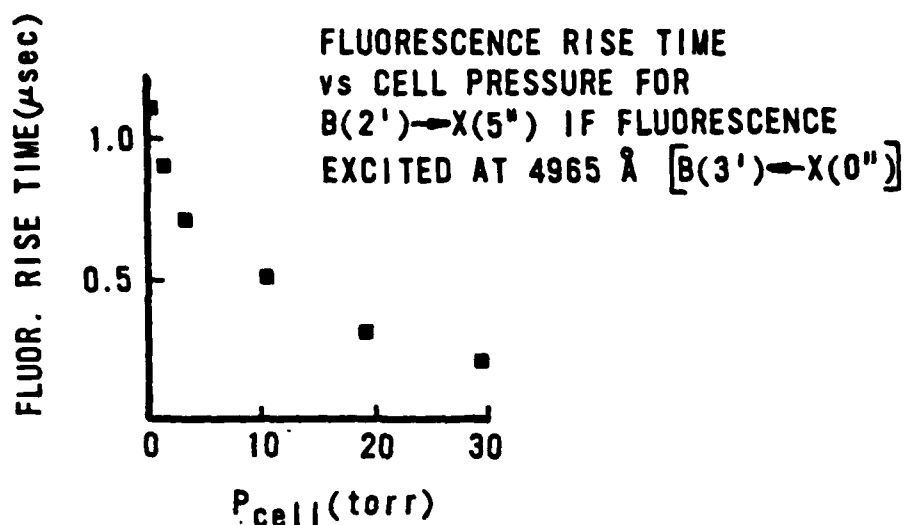
FIGURE 10 - IF B(2')-X(5") fluorescence rise time as a function of total cell pressure.

The intracavity gain detection method provides a straightforward way to make estimates of absolute gain. As discussed in the INTRODUCTION the gain "locks" the dye laser output onto the wavelength of the inverted transition when the gain is of sufficient magnitude to overcome the losses introduced by the birefringent tuning element (Lyot filter). Thus, if the loss curve of the Lyot filter is characterized this should allow us to estimate the magnitude of the gain by measuring how far it can "pull" the probe laser wavelength. The loss curve was determined by measuring the relative throughput intensity of a He-Ne laser beam as a function of the venier setting of the tuning element. Thus, if we deterimine how far the gain will pull the probe laser output in terms of venier units, we can use the curve to estimate the percent loss the gain had to overcome in order to reach threshold at the locked wavelength. Typically, the thermalized IF gain could be made to pull the probe output from 5 to 6 nm (0.23-0.28 venier units). This corresponds to a net gain of from 0.4 to 0.5%, or a gain coefficient of about 0.01%/cm for our experimental conditions.

Up to this point, we have avoided mentioning any specific problems encountered in applying the intracavity gain detection technique to a pulsed system. However, one problem in particular deserves discussion. This has to do with the existence of competing absorbing species in the test cell. In CW gain detection these losses would be compensated for by the dye gain and therefore would not be distinguised from any other steady state loss, such as mirror transmittance. However, in pulsed gain detection these absorptions give rise to probe laser intensity fluctuations which occur on the same time scale as the gain pulse, often distorting or obscuring its lineshape. For example, under the conditions favorable for gain detection on IF fluorescence it is easily seen, through intracavity absorption spectroscopy, that there is a relatively high concentration of unreacted I_2 present. I_2 absorbs in the blue, where one typically pumps IF, and predissociates into ground state atomic iodine ($^2P_{3/2}$) and excited state atomic iodine ($^2P_{1/2}$).[14,15] That such absorption occurs is easily verified through the detection of I^* fluorescence, at 1.3156 microns. Subsequent to the pump pulse, the sudden absence of the absorbing species due to dissociation causes the probe laser to turn on harder giving rise to a pulse which decays back to baseline after 1-2 msec. We presume that this is the amount of time required for fresh iodine vapor to diffuse back into the probe volume. When fluorine is added to the cell to make IF these effects are reduced in magnitude but are still present. They give rise to "false" gain pulses which occur anywhere from 15 to 100 microseconds after the intial IF gain pulse. Sometimes these pulses last for as long as 0.5 msec. Because the probe laser will shift due to the presence of a wavelength selective absorbing species, in this case I_2, its sudden absence

will also give rise to anomalous pulling effects. Initially it was supposed that the presence of these long "gain" tails may constitute evidence of chemical gain, suggesting that the I^* was acting as an energy reservoir and was pumping the IF or assisting the chemical production of IF(B). The unusually fast decay of the I^* fluorescence is most likely due to its reaction with molecular iodine to form vibrationally hot I_2 and ground state I $(k=10^{-11})$.[16] It was thought that this vibrationally hot I_2 might react through some mechanism to form some excited electronic state of IF which eventually ended up in the B state. However, because of the conspicuous absence of any side flourescence to match the "gain" tails these arguments were dismissed. The most likely explanation of the presence of the "tails" is that they are manifestations of the transient iodine concentrations in the probe volume. The details are not yet understood and further investigation is warranted.

It should be noted before continuing on to the section on CW gain detection that pulsed I_2 gain was also detected (Figure 8). I_2 was pumped at a wavelength of about 512 nm, corresponding to absorption into B(45') state. Gain was detected at 594.6 nm. It is estimated that the I_2 pressure in the cell was around 20 mtorr. The gain pulse reached maximum intensity after about 220 nsec and was relatively strong (it could be detected at this wavelength over a probe laser wavelength range of more than 12 nm). The addition of He buffer gas effectively quenched the I_2 gain. No "tails" were observed following the I_2 gain pulses. However, as mentioned above, when the pump laser was scanned to the region where the IF was pumped, anomalous pulses and tails similar to those seen in the IF gain spectra were detected.

V. CW GAIN DETECTION

The apparatus assembled for the CW gain detection experiment was similar to that shown in Figure 6. The major difference was that a CW laser was substituted for the Phase-R DL-1400 dye laser and detection system 4 was modified to include a 1/3 meter monochromator (GCA MacPherson Mo. 218). The output of the PMT coupled to this monochromator was detected using a PAR Mo. 124A lock-in amplifier, a chopper having been inserted into the pump beam of the probe laser. Also, the long pass filter was removed from the cavity since any steady-state losses due to competing absorbing species can be compensated for by supplying more gain from the dye. However, one complicating factor is introduced by removal of the filter; the gain pump laser also pumps the dye. In fact, under certain conditions it is possible to lase the probe laser by pumping the dye through the output coupler with the gain pump laser, the probe laser pump laser having been shuttered off. Thus, under normal operating conditions enhancement of the probe laser output comes not only from the gain created in the test medium but also from this extra source of dye gain. In practice, this poses no problems so long as the gain pump laser power hitting the dye jet is kept small compared to the dye laser pump power and one uses wavelength pulling as the criterion for gain detection and not just probe laser output . enhancement.

CW gain on the $B(43')-X(13")$ transition of molecular iodine was readily detected with about 20 mtorr of I_2 in the cell and pumping the $B(43')-X(0")$ I_2 absorption band with the 514.5 nm line of an Ar^+ laser (single mode power 400 mW at the head). As in the pulsed gain experiment, it is crucial to line up the pump laser so that maximum overlap of the beams is achieved. An example of the data
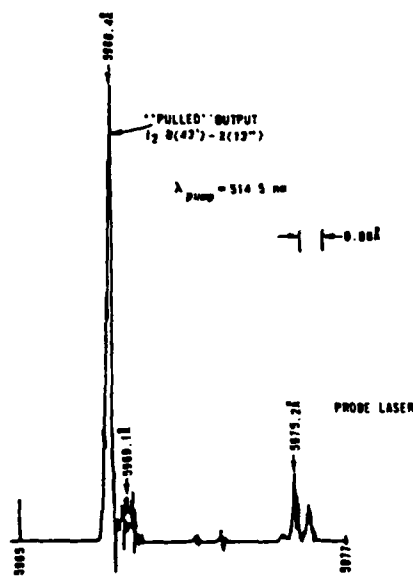
FIGURE 11 - CW I_2 gain on the B(43′)-X(13″) transition.

collected is shown in Figure 11. At these power levels the probe laser can be "pulled" approximately 1.7 nm, corresponding to a net gain of about 0.1 to 0.2%, or a gain coefficient of about 0.002%/cm for these experimental conditions. We should note that as the probe beam is tuned closer to the gain transition a greater percentage of the probe power is locked onto the transition wavelength. It should also be noted that for a given probe power a greater percentage of the probe beam can be "pulled" by increasing the I_2 pump power. Gain could be detected on the 43′-13″ transition for gain pump powers down to about 80 mW. Under these conditions the dye laser output could be pulled only a few angstroms, corresponding to a net gain of about 0.02% or a gain coefficient of about 0.0003%/cm. At gain pump powers of about 2 watts the probe laser could be pulled as far as 2.5-3.0 nm. The CW I_2 gain was effectively quenched by adding He buffer gas to the system. Gain could not be detected for buffer gas pressures above about 1.9 torr.

From the above results we estimate that the intracavity gain detection apparatus shown in Figure 6 can be used to detect net gains of around 10^{-4}, or 0.01%. These numbers agree well with those obtained by Truesdell et al.[3] who estimated the minimum detectable gain for their system based on calculated gain coefficients for the transitions considered.

VI. ATTEMPT AT DETECTING CHEMICALLY PUMPED GAIN

The experimental apparatus shown in Figure 12 was assembled in an attempt to detect chemically pumped gain using the intracavity technique described above. The apparatus was set up around the RECOIL device (a chemically pumped atomic iodine laser) located at the Chemical Laser Facility at AFWL. The flow chamber of this device was fitted with moveable Brewster windows which could be translated so that different points in the flow could be probed. Chemically pumped molecular iodine is produced in the immediate downstream vicinity of the iodine injection rake. It appears as an intense yellow-orange flame extending a few centimeters downstream from the rake. A fluorine line was added to the RECOIL device upstream from the cold trap so
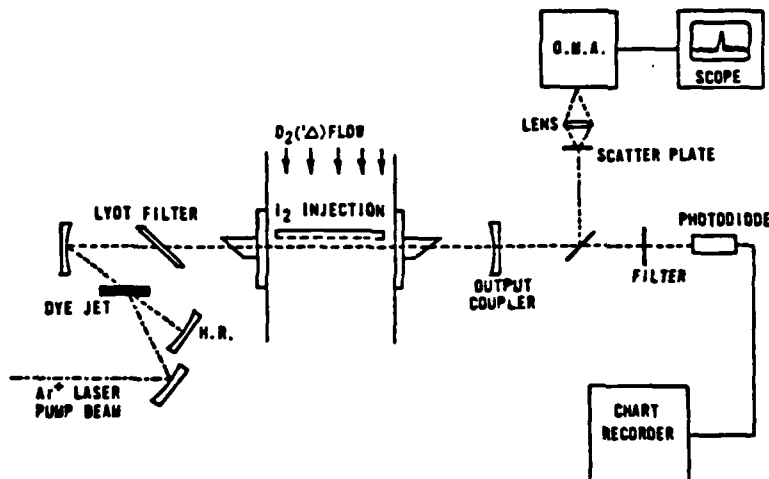
FIGURE 12 - Chemically pumped gain detection apparatus.

that chemically pumped IF could be produced in the flow. The yellow IF flame filled the entire volume of the flow chamber. The pumping agent for these experiments in $O_2(^1\Delta)$ produced in the generator of the RECOIL device. The RECOIL generator was run under conditions such that $O_2(^1\Delta)$ yields were between 20 and 35%. The pressure in the I_2 cell was held at around 30 torr giving rise to flow rates of about 6×10^{-4} moles/sec.. For those runs where IF was produced an F_2 flow rate of about 10^3 sccm was used to roughly match the iodine flow.

Figure 13 shows the results of an attempt to measure IF gain approximately 10 cm downstream from the iodine flame, at the extreme downstream edge of the reactor viewing port. The various reactor parameters are listed in the figure. The probe laser wavelength was set at 603.3 nm to overlap the IF B(0')-X(4") emission. The presence of this emission was verified earlier and both optically pumped pulsed and CW gain has been detected on this transition using the above outlined procedures. It is apparent that the probe laser is sensitive to the presence of the molecular iodine at this wavelength, however, no detectable change in the baseline
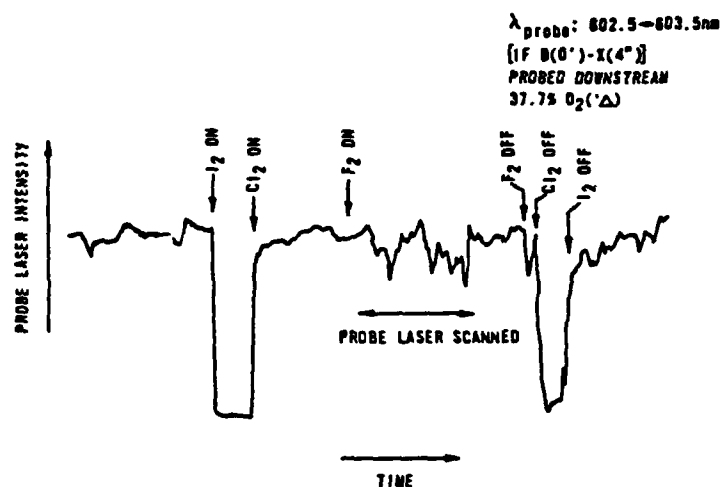
FIGURE 13 - Result of attempt to detect chemically pumped IF gain using the modified RECOIL device.

intensity is observed when fluorine is introduced into the
flow and IF emission initiated.   The baseline fluctuations
subsequent to the introduction of F_2 to the flow come
from scanning the probe back and forth over a 1 nm range and
are seen even in the absence of the F_2 flow. The IF gain
measurement was again attempted with the probe beam shifted
upstream so that it passed through the last third of the I_2
flame with virtually the same result. These results indicate
insufficient IF number density. This assertion is supported
by the fact that the IF chemiluminescence fills the entire
volume of the flow chamber, approximately 2 liters. If the
IF production and excitation could be confined to a smaller
volume relative to the probe beam volume, gain detection
would be more feasible.

Figure 14 shows the results of probing the I_2
B(9')-X(1") transition at 602.5 nm in the I_2 flame. As
expected, the transition is absorptive since there is little
probability that a population inversion can be achieved
between these two levels. However, as shown in Figure 15,
when probing the B(43')-X(13") transition at 611 nm under
nearly the same experimental conditions, the probe laser
intensity momentarily rises well above baseline, which may
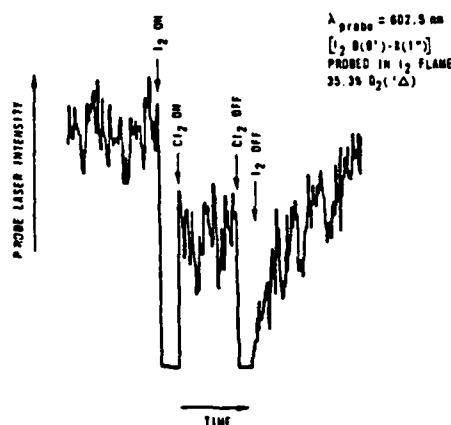
FIGURE 14 - I_2 B-X intracavity absorption spectrum taken on
chemical gain detection apparatus.
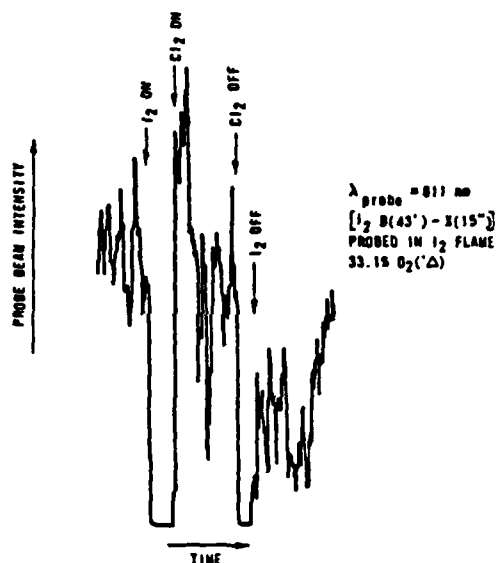
FIGURE 15 - Detection of possible gain on the B(43')-X(13") transition of I_2.

indicate gain on this transition. This feature repeated on the subsequent run, which, because of reactor conditions was the last run of the day. To date, we have not been able to get time on the RECOIL device to try to reproduce this result.

VII. RECOMMENDATIONS

First and foremost, the attempts to detect gain on a chemically pumped system should be pursued. Efforts should first be directed at reproducing the preliminary oxygen-pumped I_2^* results outlined above. Secondly, efforts should be renewed at detecting oxygen-pumped IF gain. The preliminary result, that the IF number density is too small to give rise to detectable gain or absorption, should be tested. If this is indeed found to be true, attempts to speed up the reactions and thus confine the excited IF molecules to a smaller volume could be made. For example, it is well known that the reaction of I_2 with F_2 to form IF is very slow ($1.9 \times 10^{-15} cm^2/s$).[8,17] The IF production rate

could be speeded up considerably by passing the fluorine through a microwave discharge and reacting the I_2 with F atoms.[18] This would then leave the oxygen excitation process as the rate determining step. If this reaction is too slowOr to produce sufficient number densities, other excitation sources, such as A-state nitrogen, need to be tried.

Another aspect of this project which deserves further attention is the effect of I_2 absorptions on the optically pumped IF gain pulse. In the CW mode these absorptions are compensated for by the dye laser gain. However, as mentioned above, in the pulsed mode transient I_2 absorptions exist which often occur on a time scale commensurate with the gain pulse. Thus, the IF gain pulse lineshape is often distorted, sometimes having a long tail when detected at certain wavelengths. The role of transient molecular iodine absorptions in originating this tail is not clearly understood. This role needs to be clarified.

Finally, the material presented above demonstrates the versatility and sensitivity of intracavity spectroscopic techniques. These techniques have general applicability, not only as gain diagnostics, but also in the more general study of the transient and steady-state populations of both ground and excited state species in optical resonators. Such studies should provide valuable insight and guidance in developing and troubleshooting new laser systems.

Acknowledgement

REFERENCES

1. M.B. Klein, Opt. Commun. 5, 114 (1972).

2. M.B. Klein, C.V. Shank, and A. Dienes, Opt. Commun. 7, 178 (1973)

3. K.A. Truesdell, R.A. Keller, and E.F. Zalewski, J. Chem. Phys. 73, 1117 (1980).

4. R.A. Keller, E.F. Zalewski, and N.C. Peterson, J. Opt. Soc. Am. 62, 319 (1972).

5. T.W. Hansch, A.L. Schawlow, and P.E. Toschek, IEEE J. Quantum Electron. QE-8, 802 (1972).

6. R.A. Keller, J.D. Simmons, and D.A. Jennings, J. Opt. Soc. Am. 63, 1552 (1973).

7. W. Demtroder, Laser Spectroscopy, Springer Series in Chemical Physics 5, Springer-Verlag, New York (1982).

8. P.D. Whitefield, R.F. Shea, and S.J. Davis, J. Chem. Phys. 78, 6793 (1983).

9. J.M. Herberlin, J.A. McKay, M.A. Kwok, R.H. Ueunten, D.S. Urevig, D.J. Spencer, and D.J. Benard, Appl. Opt. 19, 144 (1980).

10. M.A. Kwok, J.M. Herberlin, and R.H. Ueunten, Opt. Engineering 21, 979 (1982).

11. J. Tellinghuisen, J. Quant. Spectrosc. Radiat. Transfer 19, 144 (1980).

12. S.J. Davis and L. Hanko, Appl. Phys. Lett. 37, 692 (1980).

13. S.J. Davis, L. Hanko, and R.F. Shea, J. Chem. Phys. 78, 172 (1983).

14. J. Tellinghuisen, J. Chem. Phys. 58, 2821 (1973).

15. S.J. Davis, Appl. Phys. Lett. 32, 656 (1978).

16. G.E. Hall, W.J. Marlnell, and P.L. Houston, J. Chem. Phys. 87, 2153 (1982).

17. P.D. Whitefield and S.J. Davis, Chem. Phys. Lett. 83, 44 (1981).

18. E.H. Appelman and M.A.A. Clyne, J.C.S. Faraday I 71, 2C⁻2 (1975).

1983-1984 AFWL-SCEEE WEAPONS LABORATORY SCHOLAR PROGRAM

Sponsored by the

AIR FORCE WEAPONS LABORATORY

Conducted by the

SOUTHEASTERN CENTER FOR ELECTRICAL ENGINEERING EDUCATION

<u>FINAL</u> <u>REPORT</u>

TRANSIENT PROTECTION OF ELECTRONIC CIRCUITS

Prepared by: Ronald B. Standler

Academic Department: Electrical Engineering

University: The Pennsylvania State University

Research Location: .Air Force Weapons Laboratory

AFWL Research Contact: Major James Kee

Date: September 7, 1984

Contract Number: F49620-82-C-0035

TRANSIENT PROTECTION OF ELECTRONIC CIRCUITS

Ronald B. Standler

TRANSIENT PROTECTION OF ELECTRONIC CIRCUITS

Ronald B. Standler

3. Gas Tubes

    A. Physics of Operation

    B. Spark Gaps

        1. two electrodes

        2. three electrodes (for balanced line)

        3. coaxial mounting

    C. Neon lamps

    D. Dielectric stimulated arcing devices

4. Varistors

5. Avalanche and Zener diodes

6. Semiconductor diodes and rectifiers

    A. V,I relation

    B. Ordinary devices

        1. core switching devices

        2. rectifiers

        3. fast-recovery rectifiers

        4. Schottky barrier diodes

    C. Low-leakage diodes for electrometer inputs

        1. JFET gate to channel diode

        2. Ga As diodes (LEDs)

    D. PIN diodes for high frequency shunt applications

7. Thyristors (Shunt Components)

    1. SCRs, Triacs, ASCRs

    2. Silicon bidirectional switches

    3. Diacs

8. Impedances and Current Limiters

A. Resistors

B. Positive Temperature Coefficient (PTC) Devices

C. Inductors

    1. choke coils

    2. ferrite

    3. bifilar coils for common-mode inductance

D. Capacitors

E. Prefabricated filter modules (brief)

F. Lossy line

    1. coaxial cable with resistance wire

    2. Capcon, Inc. proprietary cable

G. Fuses and circuit breakers

    1. properties of fuses

    2. application of fuses in transient protection circuits

9. Isolation devices

A. Introduction

B. Isolation transformers (brief)

C. Optoisolators

10. Parasitic inductance

    A. Effect on clamping voltage of shunt elements

    B. Techniques to minimize parasitic inductance of shunt elements

        1. conductors of minimum length

        2. no sharp bends

    C. Four terminal construction of shunt elements

    D. Bypassing DC power supply bus

11. Applications in signal circuits

    A. Standard circuit with spark gap and avalanche diode

    B. Diode clamps to power supply bus

    C. Balanced line applications

    D. Protection of an operational amplifier

        1. inputs of inverting voltage amplifier

        2. inputs of non-inverting voltage amplifier

        2. outputs

    E. Protection of logic gates

    F. Optoisolator circuits for digital data

        1. receiver

        2. transmitter

12. Transient protection of mains (120 V_{rms} power line, 50 to 400 Hz)

    A. Introduction

    B. Secondary Arresters at point-of-entry into the building

    C. Line conditioner

    D. Uninterruptible power supply

    E. Surge protection module

    F. Surge protection inside a chassis

    G. Transient suppression at source

13. Applications in DC power supply (+5, ±15, +28 volts at 2 A or less)

    A. Introduction

    B. Protection from transients on mains

    C. Bypassing

    D. Protection of voltage regulator

    E. Protection from transients at loads

    F. Crowbar

    G. Uninterruptible DC power supply

14. Upset detection circuits

    A. Introduction

    B. Avoidance

    C. Digital circuits

    D. Coordination of transient protection and upset avoidance

    E. Transient detection on mains (120 V_{rms}, 50 to 400 Hz)

    F. Application of upset avoidance circuits in signal circuits

    G. Upset due to interruption of mains

PREFACE

The citation of trade names and names of manufacturers in this report is not to be construed as approval of commercial products, nor is the absence of a particular manufacturer's name to be construed as an adverse judgment by the author.

The schematic circuit diagrams, description of circuits, specifications, or other data in this report are not to be regarded (by implication or otherwise) as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention.

ACKNOWLEDGMENTS

CHAPTER ONE: SOURCES OF TRANSIENTS

NATURE OF ELECTRICAL OVERSTRESS PROBLEM

Because many modern semiconductor devices (small signal transistors, integrated circuits) can be damaged by potential differences that exceed about 20 to 40 volts, the survivability of modern electronics is limited. Modern electronic technology has tended to produce smaller and faster semiconductor devices, particularly high-speed digital logic, microprocessors, metal-oxide-semiconductor (MOS) memories for computers, and GaAs FETs for microwave use. This progress has lead to an increased vulnerability of modern circuits to damage by transient overvoltages, owing to the inability of small components to conduct large currents.

Smaller devices make a more economical use of area on silicon wafers and decrease components cost. These smaller devices also have less parasitic capacitance and are therefore faster. However, (in very simple terms) devices fail when the current per unit area becomes too large. The magnitude of transient currents is determined principally by external circuit parameters (e.g. the nature of the source, characteristic impedance of transmission lines, resistance and inductance between the source of the transient and vulnerable circuit, etc.) Smaller devices obviously have less area, and are thus more vulnerable to damage from a current of a given level.

Transients in electronic circuits can arise from any of several causes:

1. Lightning

2. Electromagnetic pulse produced by nuclear weapons (NEMP)

3. Electrostatic discharge (ESD), triboelectricity

4. Switching of loads in power circuits

These transients can be coupled to vulnerable circuits in two different ways:

1. Direct injection of current, for example: a lightning strike to a conductor

2. Effects of rapidly changing electric and magnetic fields, for example: induced currents from nearby lightning or NEMP.

Even without the highly vulnerable modern semiconductors, the transients in our world have an awesome capability to damage equipment. For example, a high-altitude nuclear explosion in 1962 produced an electromagnetic pulse that caused fuses in street lights to open, and burgular alarms to go off in Hawaii, which was over 800 miles away (Glasstone and Dolan, 1977, p.523).

Therefore, transient protection techniques are necessary to prevent catastrophic damage to electronic equipment from any of several different causes.

In general, protection techniques can be divided into three classes:
1. shielding and grounding
2. filters
3. non-linear transient protection devices

Shielding, while important, is not sufficient protection against electromagnetic fields from lightning or nuclear weapons because compromises in the integrity of the shield must be made (e.g. windows in aircraft, long lines must enter the shielded volume to supply electric power and carry communication signals). Variou shielding and grounding techniques are covered in detail in books by Ott (1976), Morrison (1977), and Ricketts, Bridges, and Miletta (1976) and in government reports by Lasitter and Clark (1970), and Sandia Laboratories (1972). The design of filters is covered in many electrical engineering text and reference books. The emphasis in this report is on the third class of techniques, non-linear transient protection devices. This report discusses the properties of various components that are useful for transient protection and gives examples of simple circuits that are effective in protecting representative electronic devices.

In the remainder of this chapter, we discuss various sources of transients, including laboratory simulations of "real-world" transients. This discussion is deliberately concise: interested readers will find a more detailed account in the references that are cited.

LIGHTNING

The physics of the lightning discharge have been reviewed by Uman (1969) and Uman and Kreider (1982). There are two common forms of lightning: cloud-to-ground lightning and intracloud lightning.

Cloud to ground lightning begins when a highly ionized plasma, called the

"stepped leader," propagates from a thundercloud toward the ground. When the leader is within about 50 metres of the ground, another electrical discharge, which is called a "streamer", propagates upward from the ground and establishes a highly conducting path between the ground and the stepped leader. At this time an arc current, called the "return stroke", flows from the ground up the ionized channel into the thundercloud. The return stroke produces the intense luminosity that is seen as lightning. At the end of a return stroke, a "continuing current" on the order of 100 A flows in the channel. This continuing current can have durations between a few milliseconds and half a second. After a few tens of milliseconds or more, another leader can travel down the same ionized channel toward the ground and produce a second return stroke. This process can repeat itself several times. The entire event is called a "lightning flash." One flash typically contains between three and five leader-return stroke sequences. Most lightning research has been directed at understanding the return stroke process in a cloud to ground lightning flash.

The following list summarizes some of the major parameters in cloud to ground lightning flashes.

PARAMETER	TYPICAL VALUE	WORST-CASE VALUE
Peak return stroke current, I	20 kA	$\approx$ 100 kA
Total charge transfer	20 coulombs	
Rise time of return stroke	0.2 μs (?)	
dI/dt of return stroke	10^{11} A/s	

Risetimes of the order of 1 μs are commonly reported in the older literature. These values for the risetime are probably too large, owing to inadequate bandwidth of recording devices (e.g. tape recorders, oscilloscopes) and electronic signal processing circuits. Even when oscilloscopes with adequate bandwidth were used, the sweep rate was usually set to a relatively slow rate in order to capture most of the return stroke waveform. Therefore, data on sub-microsecond risetimes could not be obtained. Recent measurements with faster electronics, rapid analog-to-digital data conversion, and storage in semiconductor digital memories has revealed rise times on the order of 0.1 μs. These data may still suffer from limited bandwidth.

The combination of a 20 kA peak current and a 0.2 μs rise time implies a value of dI/dt of 10^{11} A s^{-1}. This large value of dI/dt implies that transient protection circuits must use radio-frequency design techniques, particularly consideration of parasitic inductance and capacitance of conductors.

While the tens of kiloampere peak currents in cloud to ground lightning are certainly impressive, one should recognize that the bulk of the charge transferred by lightning flashes occurs during the continuing currents, which are usually between 50 and 500 amperes. The continuing current is responsible for much of the damage by direct strikes, including arc burns on conductors and forest fires (Fuquay, et al., 1972; Brook, et al., 1962; Williams and Brook, 1963).

Weidman and Kreider (1977) have measured the electric field from stepped leaders, and report a mean rise time on the order of 0.3 μs, and a full-width

at half maximum of about 0.5 µs. They estimate that the peak leader currents are typically between 2 kA and 8 kA when the leader is near the ground. This is of particular importance to aircraft and missiles in flight, since these vehicles may be near a leader.

In contrast to cloud to ground lightning, rather little is known about the properties of intracloud lightning. Measurements of lightning strikes to instrumented aircraft often show peak currents of a few kiloamperes (Thomas and Pitts, 1983). The presence of the aircraft probably "triggers" many lightning events that would otherwise not have occurred. Nevertheless, data obtained in this way are certainly relevant to assessments of lightning hazards to aircraft.

EMP FROM NUCLEAR WEAPONS

When a nuclear weapon is detonated a very large flux of photons (gamma rays) are produced. These photons interact with air molecules through the Compton effect to produce pairs of electrons and positive ions. These charged particles are turned by the Earth's magnetic field to produce an electromagnetic field. This field is known as an "nuclear electromagnetic pulse" (NEMP, or more commonly, EMP). The physics of the generation of nuclear electromagnetic pulse has been reviewed by Bell (1975), Longmire (1978), and by Glasstone and Dolan (1977, p.514-540).

EMP from the detonation of nuclear weapons above the atmosphere (altitude greater than 30 km, typically between 100 and 500 km in altitude) produces a particularly intense pulse that illuminates all objects on the surface of the

earth, or in the lower atmosphere of the earth, within line of sight of the weapon. A burst 300 to 500 km above Kansas would illuminate most of the United States of America (except Alaska and Hawaii). This pulse is known in military jargon as HEMP, for "high altitude nuclear electromagnetic pulse." The analytical expression for the unclassified HEMP waveform is given by Eqn. 1, where E is the electric field in units of volts per metre, t is time in seconds, and H is the magnetic field in amperes per metre (Stansberry, 1977).

$$E(t) = 5.25 \times 10^4 \ [\exp(-4 \times 10^6 \ t) - \exp(-4.76 \times 10^8 \ t)] \qquad (1)$$

$$H(t) = E(t)/377$$

It is important to recognize that Eqn. 1 is a simple model that is representative of HEMP. The details of the EMP waveform may vary depending on type of weapon, explosive yield, altitude of burst, and distance between the burst and observer. This EMP waveform given in Eqn. 1 is plotted on two different time scales in Fig. 1-1. As shown in Fig. 1-1, the EMP waveform is a pulse with a (10% to 90% of peak) risetime of about 5 nanoseconds, a duration of about a microsecond, and a peak electric field of 50 kV/m.

HEMP is a particular threat to systems that depend on electrical power or data that are carried on long lines. This includes nearly every civilian and many military systems. The peak current in overhead lines may be on the order of 10 kA (Vance, 1975; Glasstone and Dolan, 1977, p.530). Considerable shielding from high-altitude EMP can be obtained by burying the cable in the ground. However, depending on soil conductivity and depth of burial, peak currents are predicted to be between a few hundred amperes and a few thousand
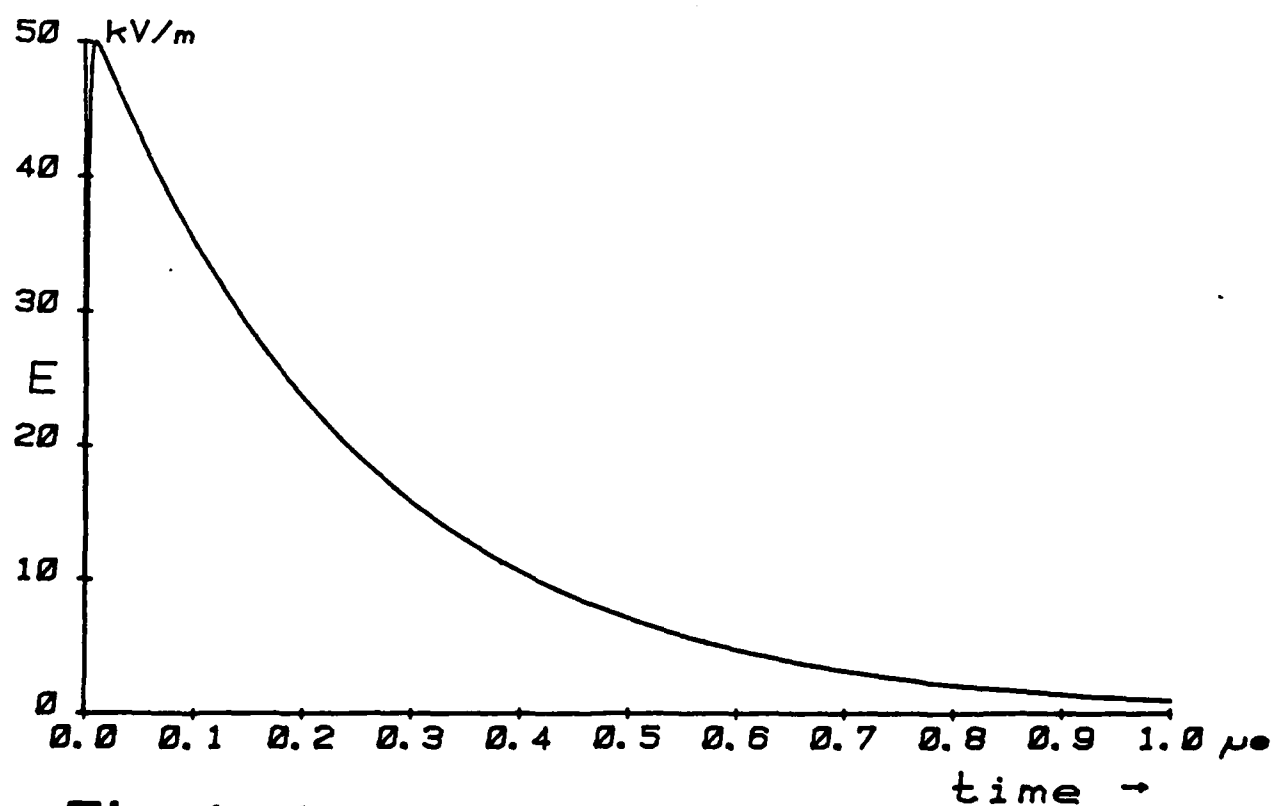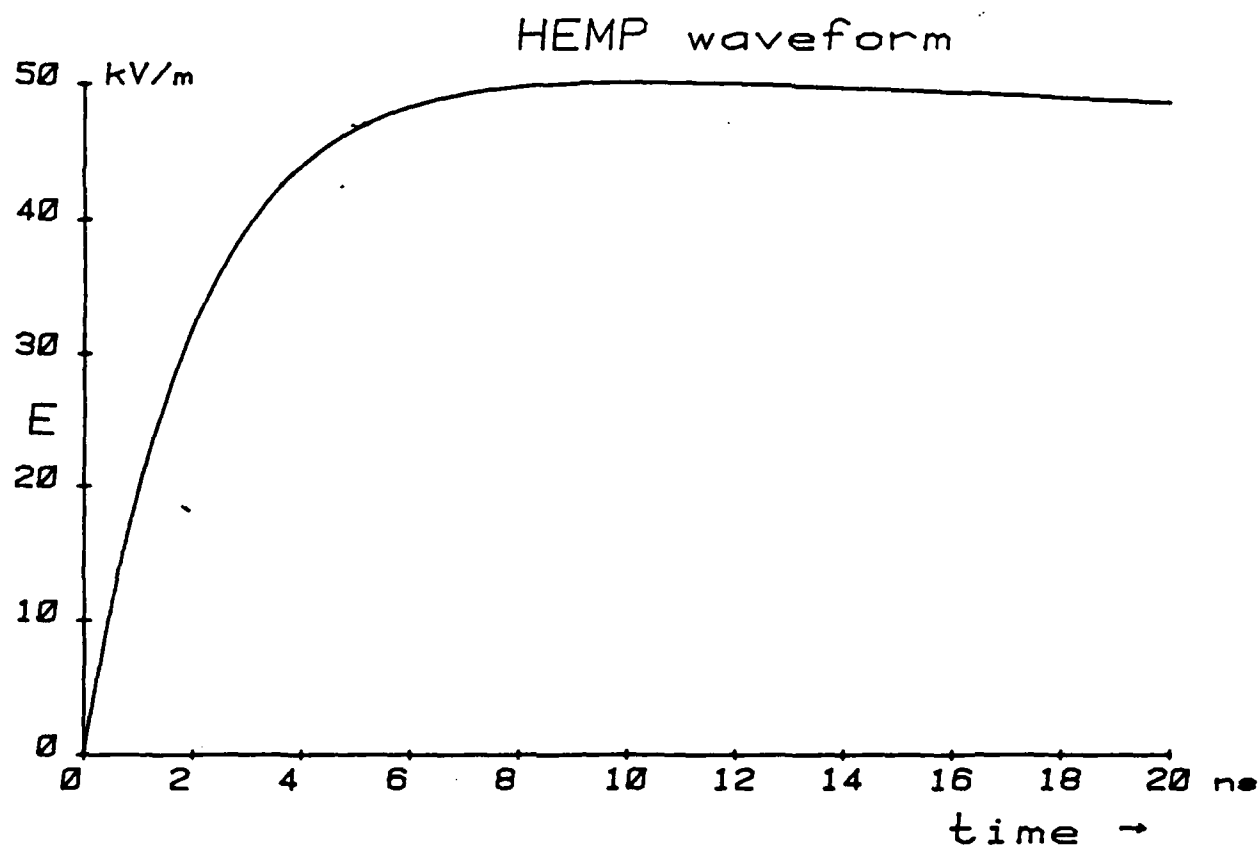
Fig. 1-1

amperes. Such currents are not negligible.

HEMP may also adversely affect aircraft and missiles in flight. The
electromagnetic field changes produce skin currents that excite resonances in
the aircraft. The resonance frequencies are generally between 1 MHz and 20
MHz. The electromagnetic field from skin currents, in turn, induces currents
in cables inside the wing and fuselage. In an independent process,
transducers and antennae on the exterior of the aircraft are illuminated
directly by the electromagnetic field from lightning or HEMP. Electronic
circuits that are connected to these transducers or antennae may be exposed to
relatively large transient currents.

The EMP from surface bursts, in contrast to high-altitude bursts
described above, is confined to a relatively small region, about 3 km to 8 km
in radius (Glasstone and Dolan, 1977, p.517-8). This region will also be
affected by blast and thermal radiation. However, surface burst EMP is still
an important problem for hardened targets (e.g. underground missile silos, and
command bunkers). Combat troops, who are far enough from ground zero to
survive the blast, may have their electronic equipment destroyed or degraded
by EMP. Furthermore, EMP from surface-bursts will produce large transient
currents on both overhead and buried cables (e.g. power lines, telephone
lines, etc.). These transient currents could travel far from the region of
the burst and cause damage to facilities that would not be affected by the
blast or heat from the burst. There is no standard waveform for surface-burst
EMP because this phenomenon is strongly dependent on the type of weapon,
explosive yield, altitude of burst, and distance between the burst and
observer. In general the electric field has a risetime of a few nanoseconds

to a positive peak, followed by a negative peak after a few tens of microseconds. Surface-burst EMP has relatively more energy at frequencies below 100 kHz than does EMP from bursts above the atmosphere.

Little unclassified information is available on the effects of EMP on modern electronic systems. Weapons tests in the 1950s and 1960s were mostly concerned with damage from blast and thermal radiation from detonations in the lower atmosphere or at the earth's surface. At long ranges, the magnitude of EMP from such tests is small, compared to that from bursts above the atmosphere. Also, electronic systems in use by the military during that time employed vacuum tubes, not semiconductor integrated circuits. We now recognize that integrated circuits are many orders of magnitude more vulnerable to damage by EMP than vacuum tubes.

Most of the unclassified data on NEMP from exo-atmospheric bursts comes from the STARFISH PRIME test, 8 July 1962 at about 23 h Hawaii local time. A 1.4 megaton device was detonated at an altitude of 400 km above Johnson Island (Glasstone and Dolan, 1977, p.45, p.523). Considerable disruption occured in civilian systems in Honolulu, Hawaii as a result of NEMP from this burst. In particular, "hundreds" of burgular alarms were activated by NEMP, and many circuit breakers or fuses were opened.

Modern systems may be much more vulnerable to damage by NEMP owing to the increasing use of integrated circuits. One anecdotal report (Science News, 30 May 1981, p.344) states that waves from an EMP simulator managed to damage electronic ignition systems in automobiles, so that the cars would not start after the tests.

ELECTROSTATIC DISCHARGE (ESD)

Electrostatic discharge (ESD) is a widely recognized hazard during shipping and handling of many semiconductor devices. Semiconductors that are especially sensitive include those that contain unprotected metal oxide semiconductor field effect transistors (MOSFETs), semiconductors for use at microwave frequencies, and very high speed logic with switching times on the order of 2 ns or less. In response to the ESD threat, most semiconductors are routinely shipped in containers that are made of conductive material. A less suitable, but commonly used, method of protection is to use plastic containers that have been treated to reduce triboelectric charging, the so-called "antistatic treatment." In addition to these shipping precautions, assembly line workers are grounded, use grounded soldering irons, ionized air blowers, and other techniques to avoid applying large potential differences to semiconductors during assembly. ESD continues to be a problem for many circuits after assembly: these completed circuit boards are shipped in conductive bags to avoid damage.

Most of the anti-ESD technology has been concerned with damage that occurs during shipping and assembly of vulnerable devices. However, ESD is still a hazard to many electronic systems during normal use. Any person who walks across a carpet and touches a keyboard of a computer terminal may damage the electronics inside the keyboard. Conventional techniques to avoid this hazard are to install conductive carpets or mats near computer terminals, or to routinely spray regular carpets with anti-static chemicals. These techniques require a conscientious and knowledgable user of equipment.

The electrical waveform involved in ESD is a brief pulse, with a rise time of a few nanoseconds and a duration of 0.1 μs to 0.3 μs (Tucker, 1968). The peak potential difference can be as large as 30 kV, but is more commonly 0.5 kV to 5 kV. If the ESD source is a charged person, the person can be modeled approximately as a 150 pF capacitor in series with a 1 kΩ resistance. This model allows us to estimate the available energy and current in ESD from charged people. If the potential difference between earth ground and the person is 1 kV, then the energy stored in the capacitor is about 75 μJ.

AC POWER FOR BUILDINGS

Buildings that are served by an overhead AC power distribution line have a major transient protection problem. The overhead line can be struck directly by lightning. The overhead line will also act as a very long antenna for electromagnetic field changes due to lightning and NEMP. These transients can be attenuated (but not eliminated) by burying the AC power line.

If lightning should strike the overhead power line outside of a building, not all of the lightning current will travel into the building (Martzloff, 1980). Part of the lightning current will travel down the line away from the building, part will flow to ground at the nearest distribution transformer, part will travel down the pole that got struck. In addition, some of the lightning current may be shunted to ground by arcs between the overhead wires and adjacent trees. Despite this current division, a lightning strike on an overhead power line can be a disaster for electrical equipment inside nearby buildings.

AC power lines also have transient overvoltages from switching of reactive loads (e.g. motors, capacitors for correction of power factor). Martzloff and Hahn (1970) found that a 1.4 kV to 2.5 kV peak transient was injected into the 120 Vrms power line when the ignition system of an oil furnace fired.

Martzloff and Hahn (1970) report that failures of electric clock motors dropped to 1% of the previous value when the manufacturer increased the insulation level from 2 kV to 6 kV. This implies that there are a considerable number of transients on 120 V rms power lines with peak voltages that are greater than 2 kV, but less than 6 kV.

In the absence of transient protection devices (e.g. lightning surge arresters, varistors), the maximum transient voltage inside a building on 120 V rms branches is limited to between 6 kV and 10 kV by insulation breakdown in the wiring and devices (e.g. outlets, circuit breaker boxes, etc.).

The cables that carry the electric power from the circuit breaker panel near the point of entry to the wall outlet will act as a transmission line with a characteristic impedance (sometimes called "surge impedance") during the initial part of the transient. Typical values of this characteristic impedance are between 100 Ω and 300 Ω (Martzloff, 1983; IEEE Std 587-1980, section 8.7). The voltage limit described in the previous paragraph and the characteristic impedance together limits the initial current to less than 100 A (e.g. 6 kV/100 Ω = 60 A). Since the initial wave travels at about 200 m/μs, and most indoor cables have a length of less than 200 metres, transmission

line concepts will govern only the first few microseconds of a transient. Thereafter, the transient current depends on the load impedance and nature of the transient, not the characteristic impedance of the cable. Peak currents of a few kiloamperes are possible on branch circuits (Martzloff, 1983).

In homes without transient protection devices on the power lines, one may hear an occasional click or pop from the stereo loudspeaker when the stereo amplifier is on. This transient usually originates when an inductive load on the AC power line (e.g. refrigerator or air-conditioner motor) is switched off. The transient propagates on the AC power line, passes through the DC power supply in the stereo amplifier, and appears at the output terminals of the stereo amplifier that are connected to the loudspeaker. While such a transient is a minor annoyance, it does illustrate two important points. First, dissimilar systems can be coupled via the AC power line. Second, transients are ubiquitous. The same transient that annoys you with a click in your loudspeakers could alter the contents of your computer's memory and cause the computer to "crash."

OTHER SOURCES

Not all transients have "exotic" sources such as lightning or nuclear weapons. Interrupting the current in an inductive load is well known to generate a high voltage pulse that can cause dielectric breakdown of insulation. A simple demonstration is to unplug a vacuum cleaner in a dark room while the motor is running. The sudden interruption of current in the motor will create an arc in air at the socket where the connection is broken.

Interruption of magnetizing current in a transformer can generate transients that are several times greater than the expected peak secondary voltages (Smith and McCormick, 1982, p.10). This can destroy rectifiers and voltage regulators in a power supply, as well as threaten vulnerable loads that are connected to the power supply.

Yet another source of electrical overstress is the accidental connection of the mains to a signal line (e.g. telephone, cable television), which is called a "power cross." Strictly speaking, a power cross is a continuous phenomenon, not a transient. However, the techniques for assuring the survival of signal electronics after a power cross are similar to techniques that are used for protection against transient overvoltages.

STANDARD WAVEFORMS IN LABORATORY TESTS

Laboratory waveforms for transient tests are commonly described in a $\alpha \times \beta$ μs format. This indicates that the waveform has a zero to peak risetime of α microseconds, followed by an exponential decay that reaches half of the peak value at β microseconds. The peak open-circuit voltage or the peak short-circuit current from the generator must also be specified. In addition, the output impedance of the generator should also be specified, especially if the output voltage was given.

The standard test waveform for avalanche diodes is a 10 × 1000 μs waveform. A 10 × 1000 μs waveform with a peak voltage of 1 kV is more severe than 99.9% of surges (presumedly from lightning) that have been observed on

paired telephone cable in Canada in 1968-1969 (Bennison, Ghazi, and Ferland, 1973). This waveform has a slower rise time by about two orders of magnitude than cloud to ground lightning. It is obvious that the transients that propagate on telephone lines have been distorted by the transmission line. High frequency componets of waveforms are attentuated by the skin effect and loading coils in combination with the shunt capacitance of the line.

Spark gaps and metal-oxide varistors are usually tested with an 8 × 20 µs waveform (Richman, 1983) when peak currents exceed about 1 kA. A more rapid decay is used for these devices, compared to avalanche diodes, due to concern about device damage from prolonged large currents. The peak currents in tests of spark gaps and large metal-oxide varistors are commonly in the 5 to 20 kA range. If the device under test were to maintain a constant 100 volts across its terminals, a 10 × 1000 µs waveform with a 10 kA peak current would deposit about 1440 joules of energy in the device. This is a very large amount of energy, hence the half-value time should be much less than 1000 µs when kiloampere peak currents are used.

Martzloff (1982, p.5) has stated that "current, not voltage, is the independent variable" in understanding transient protection. Martzloff then says that:

> "Perhaps a long history of testing insulation with voltage impulses has reinforced the erroneous concept that voltage is the given parameter. Thus overvoltage protection is really the art of offering low impedance to the flow of surge currents rather than attempting to block this flow through a high series impedance."

Conventional transient testing technique specifies the open-circuit voltage from the generator and (sometimes) the output impedance. One might
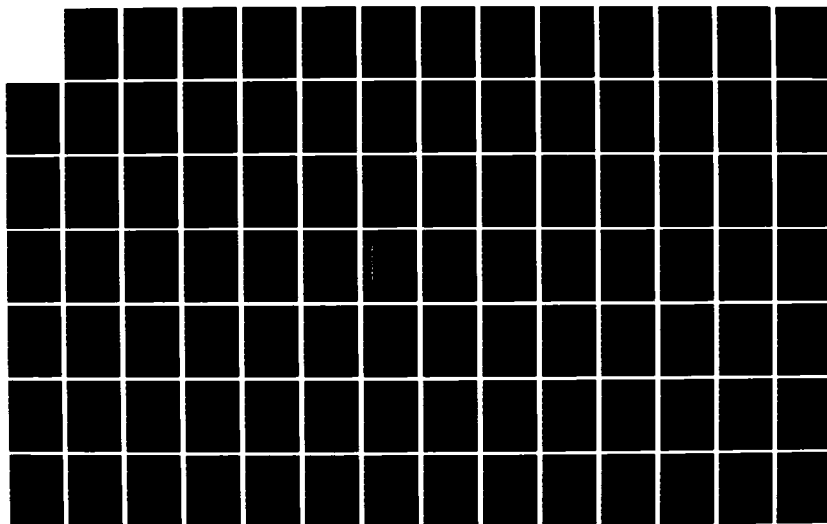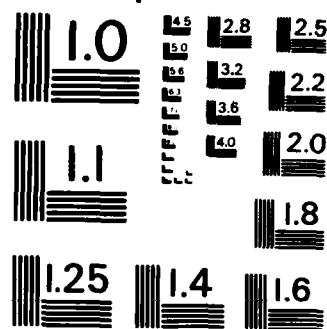
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

expect to use Thevenin's theorem to reconstruct the actual waveform. But this method is not valid, because the source often has a non-linear relation between output voltage and current. Since most of the common non-linear TPDs (spark gaps, avalanche and zener diodes, metal-oxide varistors) clamp at a nearly constant voltage, which is approximately independent of current, the open-circuit output voltage of the generator is not a very useful parameter. The short-circuit output current of the generator may be more useful, provided that the generator output impedance is much greater than the V/I value of the TPD under test.

For tests of semiconductors with currents of about 10 A or less, we prefer to use a controlled current source whose current waveform is known, and measure the voltage across the device under test. In this way, devices with very different (V,I) characteristic curves can be compared. A particularly simple test waveform is a linear ramp with a slope of about 1 A/µs. This small slope (compared to most transients) makes the effect of inductance in the package negligible. To prevent device destruction due to heating, a low duty cycle should be used (e.g. ramp for 10 µs, followed by 25 ms of zero current). For a device with a constant 100 volts across its terminals, the steady-state power input from such a test waveform would be only about 0.2 watts.

It is not necessarily a good idea to test TPDs with a waveform that resembles a worst-case transient, or even a typical transient. Laboratory characterization of TPDs should be done with a waveform that is appropriate to elicit the desired information. For example, inductance in the package can be determined from measurements taken at two different values of dI/dt that have

similar values of I and similar total energy, $\int VI\, dt$, at the two data points. Neither of these two waveforms are required to have dI/dt values that are similar to those of expected transients. As another example, one might use steady-state currents to determine the effects due to device heating, even though continuous overvoltages may not be anticipated in a specific application of the TPD.

However, proof-testing of transient protection circuits must be done with a waveform whose voltage, current, and charge transfer are all similar to the anticipated worst-case transient (even though this may require the construction of expensive surge generators). Any effort that falls short of this requirement provides little or no assurance that the transient protection circuit is adequate.

In addition to choice of waveform, there is also a question of how the test waveform is to be coupled to the system under test. All electrical connections involve two conductors, which we may call X and Y. Suppose that a two-conductor cable is cut and we call the voltage between each conductor and ground V_X and V_Y, as shown in Fig. 1-2. The common-mode, V_C, and differential-mode, V_D, representation of the voltage on the cable is given by

$$V_C = (V_X + V_Y)/2 \qquad\qquad V_D = V_X - V_Y$$

In many situations V_D represents a desirable signal and V_C is an "unwanted" voltage or noise. Use of balanced line and amplifiers with differential inputs can process V_D while rejecting V_C.

Bodle and Gresh (1961) found that only 10% of transients with peak voltages exceeding 60 volts had a differential-mode component greater than 10
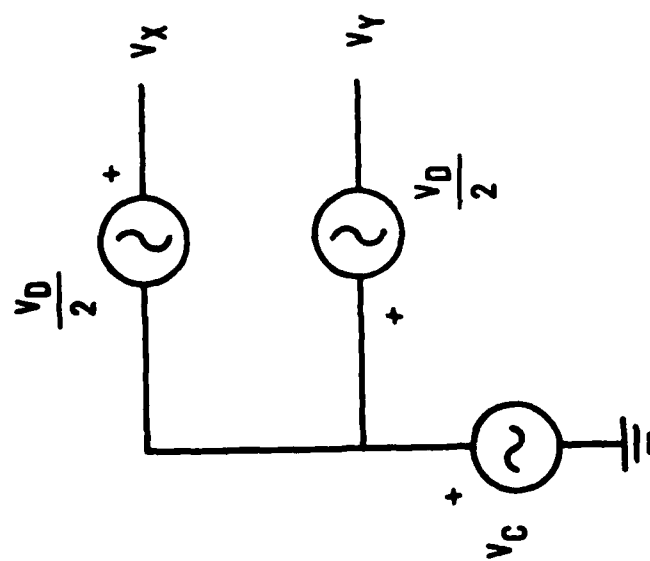
Fig.1-2

volts peak to peak. In other words, most transients were essentially common-mode phenomena. However, one might expect such a result since Bodle and Gresh studied transients on a telephone line, which is balanced. This finding has been misinterpreted to imply that all transients are common-mode phenomena and that protection from differential-mode overvoltage is unnecessary.

OVERVIEW OF PROTECTIVE CIRCUITS

There are several desirable properties of circuits for protection against electrical overstress. They are:

1. The protection circuit should have negligible effect on the system during normal operation. In particular the shunt resistance should be suitably large and both the series resistance and shunt capacitance should be suitably small.

2. The protection circuit should be fast responding: a response time of less than 1 ns is desirable if HEMP is expected.

3. The protection circuit should have good clamping: the voltage across the protected port during a large transient current should be near the maximum operating voltage of the system.

4. The protection circuit should be able to absorb worst-case transients without being destroyed.

5. The protection circuit should require minimal or no routine maintenance. Consumable components (e.g. fuses) should have indicator lamps to signal the need for replacement.

6. It would be desirable if the transient protection circuit could also protect against sustained or continuous electrical overstresses without damage to the protection circuit.

7. The protection circuit should require a small volume and be inexpensive.

Before characteristics of specific devices or design of protective circuits are discussed, it is helpful to review some general information

concerning suppression of hazardous voltage or current pulses. In this discussion, the term "signal frequency" refers to the band of frequencies that is present during the normal operation of the circuit. "Transient frequencies" refer to the range of frequencies in the overvoltage. In other words, the "signal frequency" is the "good information" and the "transient frequencies" contain the undersirable, damaging waveform.

The general form of a suppression circuit is shown in Fig. 2-1. The series impedance, Z_1, is usually a resistor. However, if the signal frequency is less than that of the transient, Z_1 may contain an inductor. If the frequency of the signal is greater than that of the transient, Z_1 may be a capacitor. In power circuits Z_1 can be the parasitic resistance of the line, as well as that of a fuse or circuit breaker.

The shunt impedance, Z_2, is usually a non-linear component; for example: a spark gap, varistor, or avalanche diode. However, Z_2 may be a capacitor or inductor that forms a low-pass or high-pass filter with Z_1, provided that the signal frequency is appreciably different from the transient frequencies. The circuit of Fig. 2-1 acts like a voltage divider. When properly designed, this circuit will provide negligible attenuation for the signal, but will remove nearly all of the transient. It is important to note that the parasitic capacitance of a varistor or avalanche diode (which can be of the order of 1000 pF) automatically forms a low-pass filter when these devices are used for Z_2 in this circuit.

Parasitic capacitance and inductance is often overlooked by circuit designers who are not familiar with radio-frequency circuit design practices.

TRANSIENT

z_1

z_2

TO
VULNERABLE
LOAD
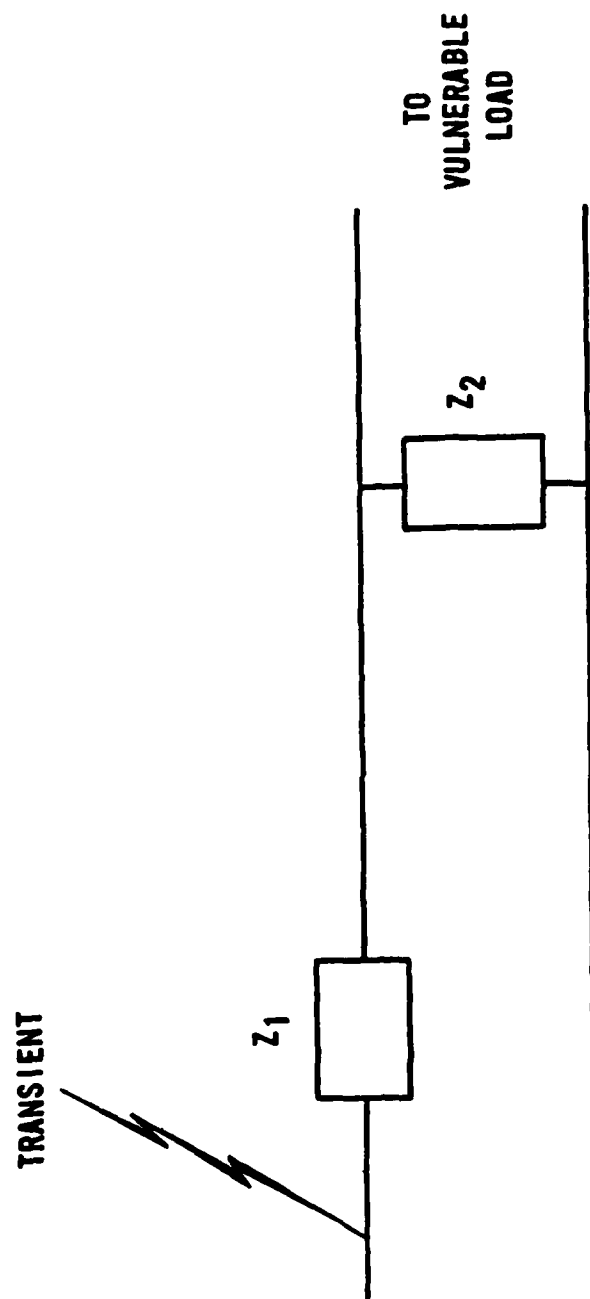
Fig.2-1

In fact, the way a component is mounted on a printed circuit board can be more important than the properties of the component itself when transients with risetimes of a few nanoseconds or less are encountered. Inductance must be minimized in the shunt path that includes components such as Z_2. Specific suggestions about how to accomplish this are given in Chapter 9 on parasitic inductance. Inductance in series with the line (e.g. parasitic inductance of Z_1) is a desirable feature, provided that the signal frequencies are sufficiently small so that the L di/dt term is negligible during normal operation.

Non-linear shunt impedances are members of a class of devices called "terminal protection devices" (TPDs) or "surge protective devices" (SPDs). A protective circuit that contains one or more TPDs is called an "electrical surge arrester" (ESA). We remark that "arrester" is often spelled "arrestor."

It is difficult to design a single stage circuit, as shown in Fig. 2-1, that can protect integrated circuits from large transients (e.g. peak currents greater than about 50 amperes). When large transients are anticipated, one usually forms a hybrid (two-stage) circuit, such as shown in Fig. 2-2. The first stage is designed to remove most of any large transient and protect the second stage. The purpose of the second stage is to protect the load. This two-stage circuit appears to have been first described by Bodle and Hays (1957), and has become the standard circuit for protection of analog and digital data lines.

We can divide transient protection circuits into three broad classes which describe how the circuit functions:
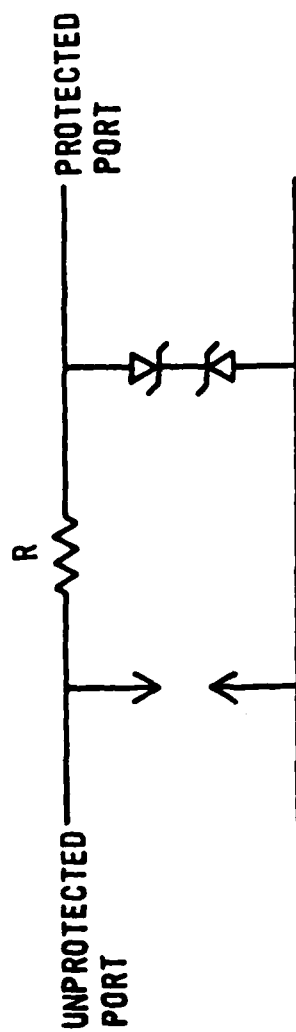
Fig.2-2

1. voltage discrimination

2. frequency discrimination

3. state discrimination.

The most common kind of transient protection circuits discriminate between normal signals and transients on the basis of voltage levels. A non-linear element, Z_2, is chosen that is essentially non-conductive at normal signal voltages, and is highly conductive at larger voltages which are encountered during transients.

When the range of signal frequencies is appreciably different from the transient frequencies, it is convenient to use frequency discrimination in the transient protection circuit. When economical filters do not provide enough attentuation of the transients, voltage discrimination circuits will also be required.

We now discuss the last class of transient protection circuits, state discrimination. There are two states: (1) the normal operation of the system and (2) the state when transients are present or anticipated. During the transient state, the protected system is instructed to ignore information that is present on the input data lines. This helps to prevent temporary malfunctions called "upsets" that arise when the transient corrupts the input data. State discrimination techniques are discussed further in Chapter 14 on upset.

DAMAGE THRESHOLD

We now consider the value of voltage, current, or power that is necessary to cause permanent damage to semiconductor devices. Such information is essential during the design of protection circuits, since the protection must attentuate overstresses to below the failure threshold.

The damage threshold is defined as the minimum power transfer through a terminal such that the device's characteristics are significantly and irreversibly altered. The damage threshold is a function of the waveshape, and is particularly sensitive to the duration of the transient.

The most widely used model for damage threshold was presented by Wunsch and Bell (1968). They showed that the maximum power, P, that could be safely dissipated in a semiconductor junction was given by

$$P = k\, t^{-1/2}$$

where t is the time duration of a pulse and k is the "damage constant." Devices with a larger value of k are able to withstand larger transients. The inverse square-root dependence on the pulse duration was derived by Wunsch and Bell (1968) for adiabatic heating of the junction. This relation is approximately valid for pulse durations that satisfy

$$0.1 \ \mu s < t < 20 \ \mu s.$$

Ideally, the value of k would be a constant for a device with a particular model number. Much effort has been devoted to finding the proper value of the damage constant, k, for hundreds of different silicon diodes and

transistors. Several conclusions are clear from this effort.

A relation of the form

$$P = A \, t^{-B}$$

where B is not necessarily one-half, fits the empirical failure data better.
Efforts to predict the values of k, A, or B from parameters on the
specification sheet (e.g. thermal resistance) have not been particularly
successful, therefore the value of k or the values of A and B must be
determined by experiment.

This simple model is known as a "thermal model" since the mechanism for
damage is melting of the semiconductor by excessive energy deposited in the
bulk semiconductor.

Enlow (1981) described variations in the mean failure threshold for
samples of 100 transistors from each of five manufacturers for four different
2N part numbers. Even for devices of the same model number and same
manufacturer, the standard deviation for the failure threshold was often about
25% of the mean value. When failure thresholds for devices of the same model
number but different manufacturers were compared, it was clear that specifying
the same model number was not adequate to assure that the two lots of devices
came from the same statistical population for failure thresholds. For
example, 2N718 transistors from Texas Instruments had a failure threshold of
452 ± 73 watts, while 2N718 transistors from IDI had a failure threshold of
97 ± 7.2 watts (these data are written as $\bar{P} \pm \sigma$). The mean for the IDI
devices is 4.86 σ from the mean of the TI devices, while the mean for the TI

devices is 49.3 σ from the mean of the IDI devices:

$$97 = 452 - (4.86 \times 73)$$

$$452 = 97 + (49.3 \times 7.2)$$

These two distributions of failure thresholds are clearly distinct.
Measurements of Texas Instruments 2N718 transistors tells nothing about IDI
2N718 transistors.

These damage thresholds are a statistical concept, not a precise number
that is applicable to a particular piece part. When damage thresholds are
determined, the device either fails or it does not fail. If it does fail, one
then has an upper bound for the damage threshold, but no information about the
effect of slightly smaller stresses. After the device fails, the experiment
can not be repeated for that particular piece part.

By testing a large number of devices one can obtain a statistical
distribution of damage thresholds and fit various models to these data. Such
effort is expensive and the results are not applicable to components of the
same model number from a different manufacturer. There are apparently no
discussions in the literature about whether such statistical distributions are
applicable to different production lots of the same model number and same
manufacturer.

One of the major reasons for this variation in damage thresholds is
inherent in the device specifications. Specifications for electronic
components give maximum or minimum values for various parameters that are
important in the original application of the device, but do not specify how
the device is to be constructed. Therefore, parts from different

manufacturers with the same model number will probably have different electrode geometries and different compositions. Also, manufacturers will revise their process for a particular model from time to time without changing the model number of the parts. Such unspecified features can be important in determining transient performance and damage thresholds.

The simple thermal model for damage in semiconductors ignores effects caused by different failure mechanisms such as:

1. second breakdown

2. metalization failure

3. breakdown of gate oxide in MOSFETs.

Chowdhuri (1965) showed that diodes that were conducting when a transient voltage was applied had a breakdown voltage that was about half of the breakdown voltage for a diode that had no initial bias. This would be expected to affect the failure thresholds for these devices. However, most empirical studies of failure thresholds are made with an initially unbiased device. These thresholds are not necessarily valid for devices that are conducting when the transient occurs. This is an important point, since during normal operation of a system many vulnerable devices may be conducting when the transient occurs, and thus have different failure thresholds.

Moreover, most empirical studies of failure thresholds of transistors apply the transient pulse to the base-emitter junction (with a polarity that will reverse bias this junction) while the collector terminal is an open circuit and there is no initial current in any of the transistor's terminals. This is certainly not the usual way to operate a transistor, and one would

expect that failure thresholds obtained in this way would not be representative of failure thresholds during normal operation.

Devices that are initially conducting may enter the "second breakdown" region of operation as a result of transient overstress. Once the transient pushes the device into second breakdown, the system power supply may kill the semiconductor. In this scenario, the transient does not necessarily need a large energy content in order for the device to fail, since the transient only initiates the failure process.

One attack on the problem of determining damage thresholds involves more research in pure and applied solid-state physics regarding failure mechanisms in devices. With more knowledge one might be able to (1) design devices with greater damage thresholds or (2) control production parameters to eliminate devices with low damage thresholds without testing every piece part. While such research is certainly worthwhile, it is of no help to today's circuit designer.

Another attack on the problem of determining damage thresholds is to use the "absolute maximum ratings" given in the specification sheet. These parameters are usually steady-state ratings. It is well-known that components can tolerate, for a few microseconds, values of current and power that are factors of 10^2 (or more) greater than their maximum steady-state ratings (Wunsch, Bell, 1968; Alexander, et al., 1975). Therefore this approach is very conservative. If protecting a circuit to the absolute maximum ratings imposes a considerable hardship on the designer, stresses of a factor of two above these steady-state maximum ratings are easily tolerable for a suitably

brief time (e.g. a few microseconds). Such an approach also avoids the problem of specifying in advance the transient waveforms to which the component will be exposed, although it will be necessary to have an estimate of the worst-case peak current and a few other parameters.

Such an approach has been endorsed by Military Handbook 419 (p. 1-50, 1982). In the absence of data from manufacturers or laboratory testing, the following surge withstand levels are given as typical:

Integrated circuits: 1.5 times normal rated junction and supply voltage

Discrete transistors: 2 times normal rated junction voltage

Diodes: 1.5 times peak inverse voltage

Under this approach, the circuit should be protected against the following conditions:

1. Rise time on the order of few nanoseconds (NEMP and ESD)

2. Continuation of stress for 0.1 to 0.5 seconds (continuing current in lightning)

3. Polarity reversals.

4. Estimate of worst-case peak current and total energy in transient.

The objection will be made by some that using the absolute maximum ratings as an estimate of the damage threshold is much too conservative since components are known to withstand greater stresses for brief periods of time. However, it is often possible to design economical protective circuits that can protect components from stresses that are greater than the absolute maximum ratings. The use of these circuits can provide substantial assurance

that equipment will survive exposure to adverse electrical environments.

UPSET THRESHOLD

Nearly all transient protection circuits allow a small fraction of the incident transient, called the "remnant," to propagate to the protected devices. In a properly designed protection circuit the remnant will have insufficient energy, current, or voltage to damage protected devices. However, there is still concern that the remnant could be misinterpreted as valid data. Such misinterpretation is called "upset." The threshold for upset is usually within the normal range of input voltages to the system. Therefore, one can not discriminate against upset on the basis of voltage levels alone.

We shall discuss some ways of dealing with the problem of upset in the last chapter in this report.

TRANSIENT PROTECTIVE DEVICES

In the next seven chapters we shall describe the properties of various components that are useful in transient protection circuits. Here we present a very brief overview of several of the major families of transient protection components.

FAMILY   ADVANTAGES AND DISADVANTAGES

spark gaps: can safely conduct large currents (5 kA for 50 μs)

                   low voltage in arc mode

                   small parasitic capacitance (<2 pF)

                   requires large voltage (≥100 V) to conduct

                   can be slow to conduct

                   possible "follow-current" (sustained short-circuit)

metal oxide varistors:

                   fast response (< 0.5 ns)

                   large energy absorption

                   can safely conduct large currents (1 kA for 20μs)

                   inexpensive

                   large parasitic capacitance (1 to 10 nF)

                   clamping voltage difficult to predict

avalanche diodes: fast response (<0.1 ns)

                   various precisely pre-determined clamping voltages

                          (between about 6.8 and 200 volts)

                   small maximum allowable current (≤100 A for 100μs)

                   large parasitic capacitance (1 nF)

switching and rectifier silicon diodes:

                   small clamping voltage (0.7 to 2 V)

                   inexpensive

SCRs and triacs: small clamping voltage (0.7 to 2 V)

can tolerate sustained large currents

slow to turn-on or turn-off (2 µs)

sustained conduction

optoisolators: large isolation voltages (5 kV)

good common-mode rejection

easy to use to receive data

difficult to use to transmit data

fast devices (<1 µs switching time) are expensive

GAS TUBES

Gas tubes are the principal component for shunting large transients away from vulnerable circuits. Typical spark gaps in a ceramic case can conduct transient current pulses of 5 to 20 kA for 10 µs without appreciable damage to the spark gap. Gas tubes have the smallest capacitance of all presently known non-linear transient protection devices. The typical shunt capacitance of a spark gap is between 0.5 pF and 2 pF. Thus spark gaps are one of a few non-linear devices that can be used to protect circuits in which the signal frequency is greater than 50 MHz. Gas tubes have three major disadvantages: (1) they can be slow to conduct, (2) in some situations they are difficult to turn off after the transient has ended, and (3) all gas tubes require at least 60 to 100 volts across the tube before the gas will conduct.

The operation and construction of spark gaps has been described by Kawiecki (1971, 1974), Cohen, Eppes, and Fisher (1972), and Bazarian (1980).

The spark gap is inherently a bipolar device; it makes no difference which electrode has positive charge. Therefore, one only needs to discuss one quadrant of the (V,I) curve to understand the behavior of the spark gap. However, some gas tubes (particularly low-voltage neon indicator lamps or DC voltage regulator tubes) have small differences in behavior depending on the polarity of the voltage across the tube. These differences are owing to a coating of barium or strontium on the electrode that is intended to be the cathode. We shall not discuss such details further because they are irrelevant to transient protection.

The (V,I) characteristic for a representative low-voltage neon-filled spark gap is shown in Fig. 3-1. Notice that, unlike conventional characteristic curves, the vertical axis is the logarithm of current. A logarithmic axis is necesary to display behavior for currents that span seven orders of magnitude. The characteristic curve is rather complicated. Each segment of the curve is associated with a particular physical process.

We shall follow the characteristic curve in Fig. 3-1 for a slowly increasing potential difference across the gap, starting at point A. At point A the gap switches from an insulating to a conducting state. The potential at A is called the "DC firing voltage of the gap."

Between A and B the incremental resistance, dV/dI, is negative; thus this portion of the curve is called the "negative resistance region." If the gap is operated in this region (e.g. in series with a 1 MΩ resistor and a 100 V DC voltage source) the discharge will flicker. This can produce intermittent noise pulses.

The portion of the curve between points B and C is known as the "normal glow region." Voltage regulator tubes are operated in this region since the voltage across the tube is approximately independent of the current. The light that is emitted from the tube comes from a "cathode glow," a thin region of excited gas atoms that cover part of the cathode surface. The area of the cathode glow is approximately proportional to the current.

When the glow covers nearly the entire cathode surface and the current
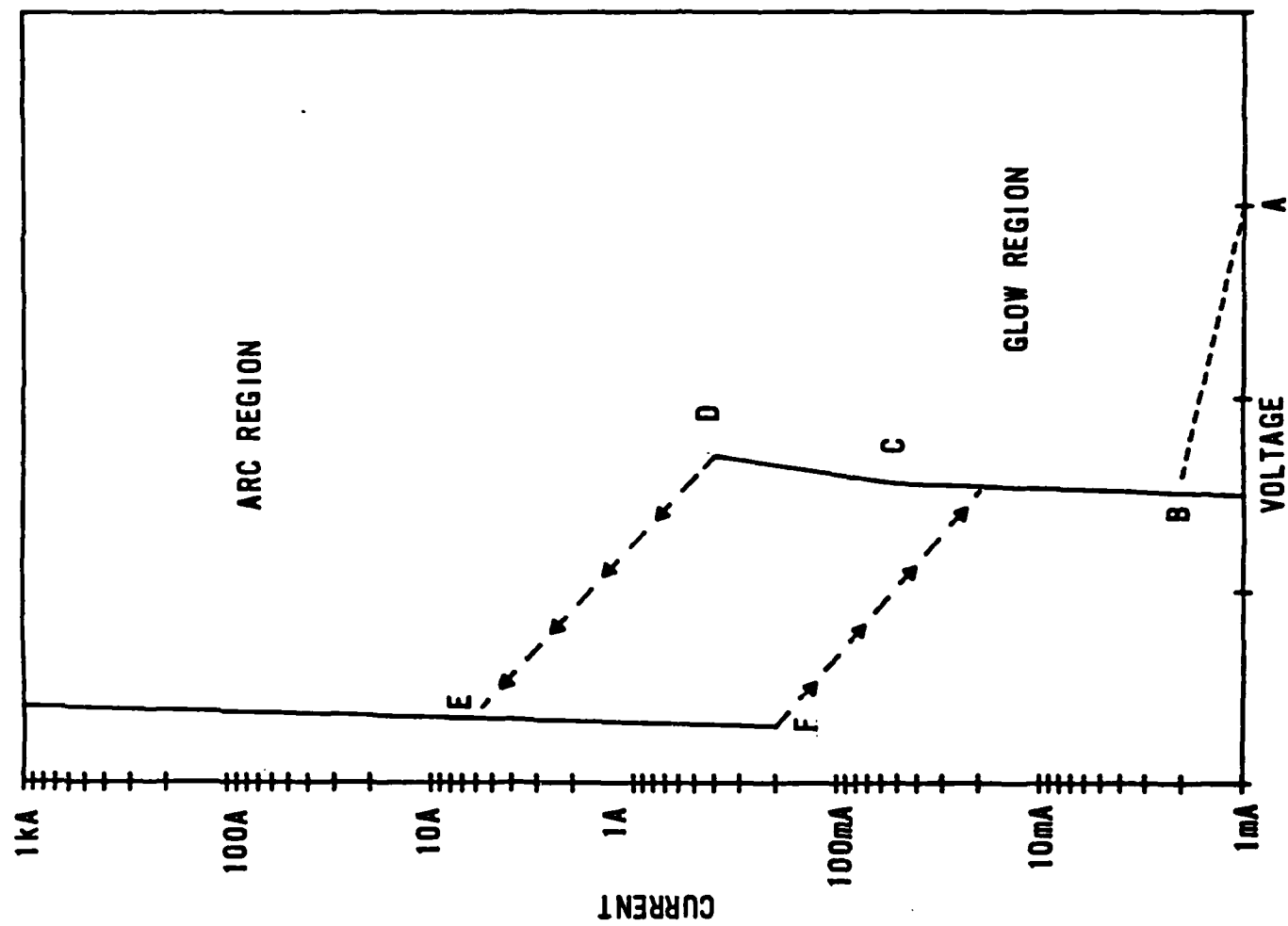
Fig.3-1

increases, the discharge enters the "abnormal glow regime." This is the portion of the characteristic curve between C and D, where dV/dI is positive.

Electron-ion pairs are accelerated in the intense electric fields which exist between the electrodes when the gas tube is operated in either the normal or abnormal glow region. When the charged particles obtain sufficient energy they can form additional ion pairs by collision with either neutral atoms or ions. This process is called an "avalanche." The avalanche is limited by the formation of space charge layers and by recombination of electron-ion pairs. The light from the discharge in either the normal or abnormal glow region contains spectral lines that are characteristic of the atoms of gas and ions. For neon the color is orange-red; for argon it is blue-violet.

When the current is increased to a sufficiently large value, point E, the discharge abruptly enters the "arc regime." In the arc regime the light comes from excited metal atoms that have been blasted from the cathode by positive ion bombardment. Positive ions are accelerated in the electric field between the electrodes and colide with the cathode in both the abnormal glow and arc regimes. The cathode material is blasted away by this bombardment, a process which is known as "sputtering." This is the source of black deposits which are seen on the walls of old neon lamps and spark gaps with glass cases. At greater currents the cathode material is removed more rapidly. The electrodes are erroded and consumed by this process. One result is that the gap spacing (and hence the DC firing voltage) increases after appreciable sputtering has occurred and the tube may no longer perform satisfactorily. Therefore, gas tubes are consumable components.

For a tube with nickel electrodes, the light from the arc appears blue. The potential across the tube is essentially constant at about 20 volts (independent of current) when the tube is in the arc regime. In transient protection applications it is highly desirable to operate the tube in the arc regime, because the voltage is clamped at the relatively small value of about 20 volts. For tubes with different gases and at different pressures the arc voltage can be from 10 to 30 volts. For typical spark gaps (with DC firing voltages between 90 and 300 volts), the maximum current in the glow region is usually between 0.2 A and 1.5 A. The value of this glow to arc transistion current is not precisely reproducible, even for the same gap.

When the current is reduced to the "arc extinguishing current," shown as point F in Fig. 3-1, the arc stops and is replaced by a glow discharge. The value of the extinguishing current is usually between 0.1 A and 0.5 A. Notice that there is hysteresis in Fig. 3-1: the current to initiate an arc is greater than the minimum current that will sustain the arc. Therefore, each of the two paths between the arc and glow regions in Fig. 3-1 can only be traversed in one direction.

When a tube is operated in the arc regime, the gas and metal electrodes become very hot. This heating can cause failure of the seal between the insulating case (glass or ceramic) and the electrodes. Alternately, the large internal pressure of the hot gas can shatter the case. A ceramic case can probably withstand higher temperature and pressure than a glass case. Many spark gaps use Kovar^{TM} (an iron-nickel alloy) sealed to an alumina ceramic case. The thermal expansion coefficient of these two materials is nearly the

same, which reduces mechanical stress on the seal during operation. The electrode metal is attached to the Kovar seal.

If atmospheric air at near sea level pressure leaks into the gas tube, the DC firing voltage of the tube will increase by approximately a factor of 10 to 20. In many situations this will destroy the tube's protective function. If the spark gap shatters or explodes, of course, the protective function is lost. Some spark gaps have a metal with a relatively low melting point inside the gap that will cause the spark gap to fail as a short-circuit if the gap is operated in the arc regime for a prolonged period (e.g. I Δt ≈ 20 to 50 ampere seconds). Such spark gaps are identified by the manufacturer as being "fail-safe."

If the normal signal source can maintain the arc discharge, then the gas will not return to the non-conducting state after the transient is completed. This condition, which is known as "power follow" or "follow current," can ruin the spark gap, and damage the wires and the source of follow current. It is important to stress that follow current can not occur if: (1) the maximum source output current and voltage are less than the arc extinguishing current and voltage of the tube (shown as point F in Fig. 3-1) or (2) if the maximum source output voltage is less than the arc extinguishing voltage (point F in Fig. 3-1). If the maximum output current of the source is less than 50 mA or the maximum magnitude of the output voltage is less than about 15 volts, then follow current is impossible in typical spark gaps. On the other hand, if a source can furnish 500 mA (or more) at 20 volts DC (or more), then follow current is quite possible.

One should not rely on the periodic zero-crossings of AC power lines to extinguish an arc. It is true that the magnitude of the voltage on a 120 V rms 60 Hz power line is less than 20 volts for 0.63 ms during each zero-crossing. However, thermionic emission of electrons from hot electrodes may sustain the arc during these brief zero-crossings.

When a spark gap is connected across a power supply (or other circuit) that can furnish follow current, one must prevent follow current from occuring. One way to do this, which is shown in Fig. 3-2, is to connect a varistor in series with the spark gap. The varistor characteristics should be chosen so that, as a general rule, there would be less than 0.1 mA of current in the varistor if the varistor alone were connected across the line during normal operation. A fuse or circuit-breaker may be included between the spark gap and source as an additional precaution.

A power resistor in series with a spark gap, as shown in Fig. 3-3, is often suggested to prevent follow current. While this technique can prevent follow current, it also greatly increases the clamping voltage owing to the voltage drop across the resistor. The IR drop can be quite substantial; for example, consider $R = 1 \, \Omega$ and $I = 5000$ A. In many situations the value of R needs to be much greater than $1 \, \Omega$, a condition that is even less favorable to protecting a circuit from transient overvoltages. Including a resistor (or an inductor) in series with a spark gap is definitely not a recommended technique.

Rovere (1934) and Poindexter (1974) have described the circuit shown in Fig. 3-4 that uses a relay to interrupt follow current. This circuit will
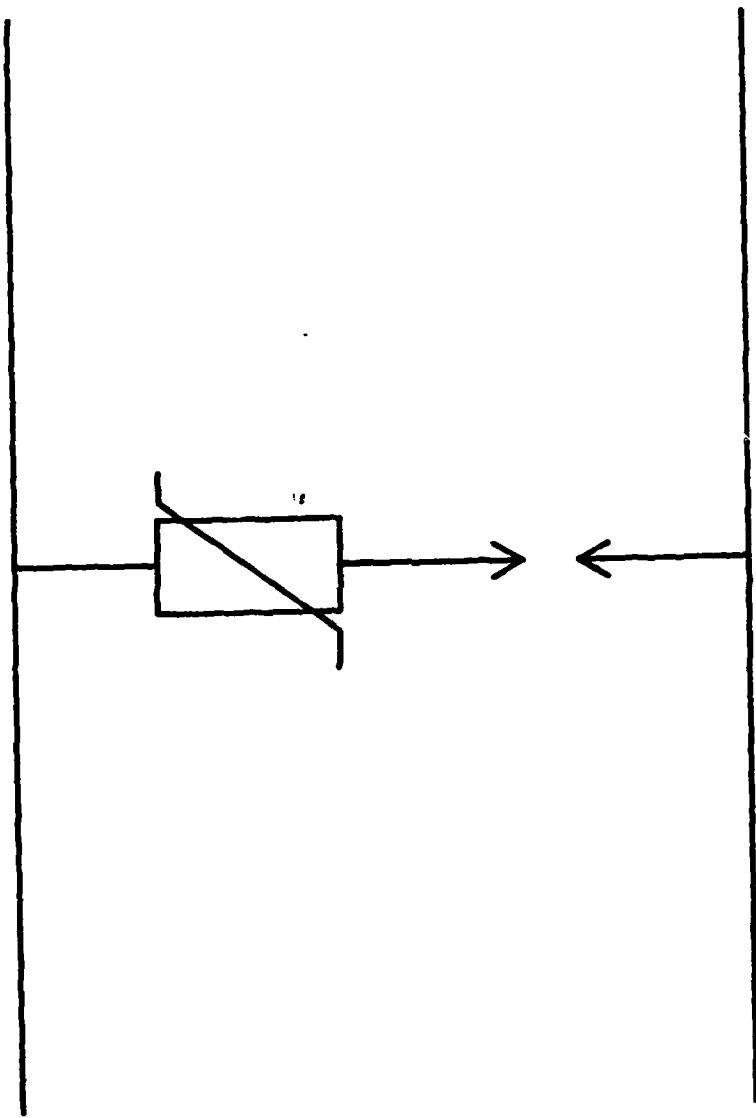
Fig.3-2
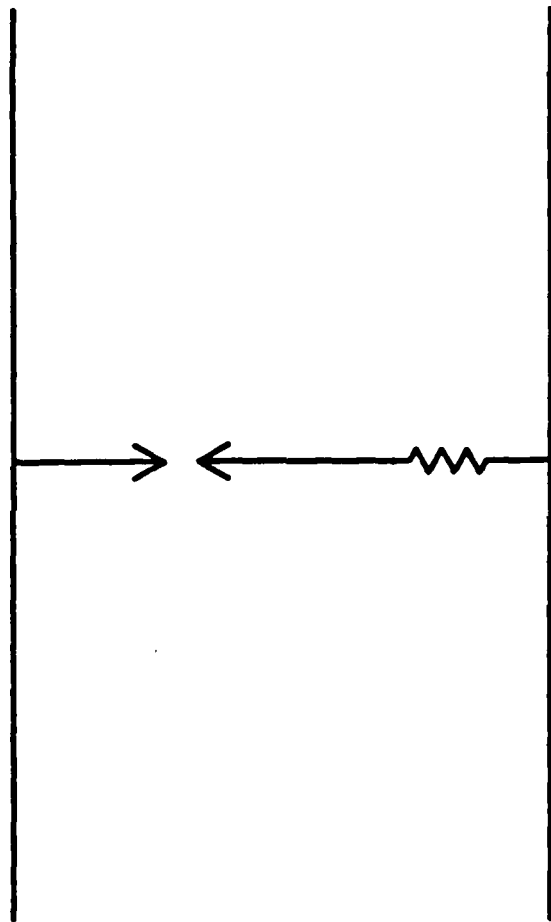
Fig.3-3

extinguish the arc by placing a short-circuit across the gas tube. However, it is a disaster for the system, since the "follow current" now flows through a short-circuit instead of the gas tube. If the "follow current" is not interrupted by a fuse or circuit breaker, the source of the "follow current" may be damaged. In any event, the power to the protected loads will be interrupted until the fuse or circuit breaker is restored. This is <u>not</u> a recommended circuit.

We can modify the circuit in Fig. 3-4 by moving the relay contacts upstream in series with both the spark gap and load, and by using normally closed contacts. That this modified circuit also interrupts the mains power to the load may be viewed as a serious disadvantage. However this disadvantage is really owing to the arc in the gas tube, and not the method of extinguishing the follow current. When the gas tube is operating in the arc mode, there is about 20 volts across the load, essentially independent of the usual mains voltage. Therefore, the arc discharge has already interrrupted the AC power to the load.

M^C Neill (1921) described a way to mount spark gap electrodes on bimetallic strips. The heat from a power-cross or follow current would cause the electrodes to move and touch each other. This was claimed to eliminate the fire hazard from heat generated by the arc. This basic idea might be applied in a different way to have the hot electrodes move farther apart. The increased gap spacing might extinguish the arc follow-current. Of course, this modified method offers no transient protection until the electrodes cool and return to their normal position.
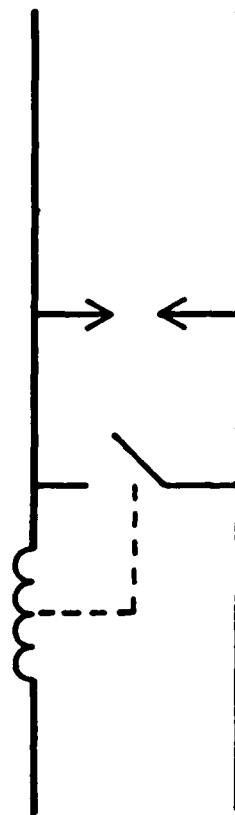
Fig.3-4   ROVERE, 1934 AND POINDEXTER, 1974

A spark gap should be selected that will not conduct during normal operation of the system. This is done by requiring that

$$1.2 \times \max|V_{normal}| \geq \min(V_{DC})$$

where $\max|V_{normal}|$ is the peak voltage during normal operation, and $\min(V_{DC})$ is the minimum DC firing voltage of the spark gap. The effects of tolerance on V_{DC} of the spark gap should be included in $\min(V_{DC})$.

There are several suggestions for increasing the reliability of circuits that are protected by spark gaps. E. Popp (1968) put two spark gaps in parallel for redundancy: if one shattered, the other would provide protection. Palmer (1977, p.28) and Sherwood (1977, p.35) both advise that the spark gap with the smallest tolerable DC firing voltage is not necessarily the best choice. They both choose 230 volt gaps for protection of overhead cable television coaxial lines, instead of 90 volt gaps. In their situation, mains voltages could be injected into the coaxial line. In effect, these engineers have defined the mains voltages as "normal" on the television cable, so that the spark gap will not conduct during a power-cross.

We now consider the transient operation of a gas tube. For rapidly changing waveforms (e.g. 1 kV/µs or more) the actual firing voltage is observed to be several times greater than the DC firing voltage (Trybus, et al., 1979). The DC firing voltage alone is sufficient to produce an electric field between the electrodes that is necessary to get ions and electrons to energies that will produce ionization by collision (a glow or arc discharge). So why does the tube require larger voltages to fire? The time delay comes from two effects: (1) some electron-ion pairs must be created in the gas by random processes and (2) a finite time is required for these initial charged

particles to form an avalanche process. These effects are called the "statistical" and "formative" times, respectively.

The initial creation of electron-ion pairs is made more frequent by introducing a beta-emitting radioactive isotope into the gas tube. Common choices are tritium gas (^{3}H), radioactive krypton gas (^{85}Kr), or radioactive nickel electrodes (^{63}Ni). The initial rate of decay is usually between 0.01 and 5.0 μCi; about 0.1 μCi is typical. (one curie [Ci] = 3.7 × 10^{10} disintegrations per second) Although one beta particle (electron) is emitted, on the average, every 0.3 milliseconds from a 0.1 μCi source, the secondary electron-ion pairs produced by the collision of the beta particle with atoms of the gas will persist for a much longer time. Thus the beta emitting isotope provides, indirectly, a source of ions that is continually present. This eliminates the "statistical" time delay.

The "formative" time for state-of-the-art spark gaps is between about 1 and 10 ns. The response time for spark gaps is commonly measured with a voltage source that has a constant rate of rise, dV/dt. For rates of rise larger than about 500 V/μs, the firing voltage of the spark gap is appreciably greater than the DC firing voltage. The response time, Δt, is given by

$$\Delta t = V_f \; / \; (dV/dt)$$

where V_f is the measured firing voltage on a waveform with a constant rate of of rise, dV/dt. The response time approaches an asymptotic limit that is an estimate of the "formative" time for spark gaps that contain radioactive material. Singletary and Hasdal (1971) found minimum response times between about 1.8 and 3.0 ns for a dozen different spark gaps with DC firing voltages between 90 and 800 volts. They noted that these times are approximately the

transit times of positive ions between the electrodes.

Various relationships exist between the firing voltage, V, and response time, Δt, of a spark gap. Joslyn Electronic Systems published a graph that describes the response time of their 2029-15 spark gaps, which have a DC firing voltage of 150 volts. The graph shows the following relation for V in volts and Δt in seconds:

$$\Delta t = \frac{6.2 \times 10^8}{(V-150)^{5.67}} \qquad \text{and } 4 \times 10^{-8} \text{ s} \leq \Delta t \leq 10^{-3} \text{ s}$$

In recent years, several companies have developed spark gaps that will conduct in less than 1 ns when a voltage is applied with a rate of change of at least 1 kV/ns.

Several different spark gap electrodes were examined by X-ray fluorescence to identify the electrode materials. The results were:

Spark Gap Manufacturer and Model	Material
Cerberus UC90	Ti in center; Fe, Ni elsewhere
Joslyn 2027-09	Fe, Ni, Ti, K, I
Siemens T61-C350	Cu
Siemens B1-C75	Mo, Ni, Fe, Ba
Genalex 39/260	Fe

The potassium in the Joslyn 2027 spark gap is due to potassium chloride that is applied to the electrode surface to provide electrons from photoionization.

The following companies, which are listed in alphabetical order, sell

spark gaps in the U.S.A.

C. P. Clare

Cook Electric, division of Northern Telecom

E. G. & G., Inc.

Genalex (distributor for M O Valve Company of England)

Joslyn

Lightning Protection Corporation

Lumex (distributor for Cerberus of Switzerland and Elevam of Japan)

Mitsubishi

Siemens

THREE ELECTRODE SPARK GAPS

There are a few models of spark gaps available that have three electrodes, as shown in Fig. 3-5. In the usual application, one connects the middle electrode to local ground (earth) and the other two electrodes to a pair of signal conductors. The three electrode spark gap is particularly useful for the protection of a balanced transmission line (e.g. telephone signals, RS-422 computer data communictions). The middle electrode of these spark gaps has a small hole that allows plasma (highly conducting gas) from one chamber to reach the other chamber within 0.1 μs. The three electrode spark gap provides both differential- and common-mode protection in a single component. Similar protection would require three separate two-electrode spark gaps, but there is no simple way to get separate spark gaps to conduct essentially simultaneously. The three-electrode spark gap provides superior
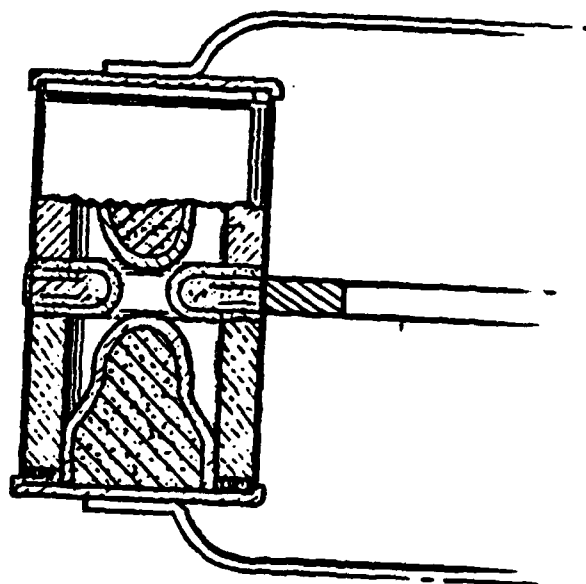
Fig.3-5

THREE ELECTRODE SPARK GAP
KAWIECKI, 1971

protection for a balanced line and lower component and assembly costs.

Prior to the introduction of three electrode spark gaps, it was common practice to connect one two-electrode gap between each line and ground. Two two-electrode gaps were used to protect a balanced pair of lines. Bodle and Gresh (1961) found that large differential-mode transients could be produced when one two-electrode gap (but not the other) fired. Most of the transients were essentially common-mode, unless unsymmetrical firing of a pair of spark gaps occurred. In this way, use of the wrong type of transient protection components (i.e. two-electrode gaps instead of three-electrode gaps) introduced a new threat to the equipment.

Three-electrode spark gaps are available from the previously cited companies that manufacture two-electrode gaps.

COAXIAL SPARK GAP

Since antennae are usually situated in elevated, exposed locations, they are commonly struck by lightning. Moreover, they can collect appreciable electromagnetic radiation from nearby lightning or EMP from nuclear weapons. It is therefore desirable to protect the electronic circuits that are connected to an antenna.

The most common protective device for antenna lines is a spark gap in a coaxial mounting. The spark gap has a very small shunt capacitance and large resistance in the non-conducting state, which gives the spark gap a small insertion loss for frequencies up to about 500 MHz. In addition to these

desirable passive properties, the spark gap excels at conducting large peak currents that would be encountered in a direct lightning strike to the antenna.

In the case of a transmitting antenna, the protective circuit must be able to withstand the relatively large voltages that are encountered in routine operation. For example, a transmitter power of 50 watts requires a peak voltage of about 90 volts (assuming sinusoidal waveform and a perfect match between the 75 Ω transmission line and a dipole antenna). Larger transmitter power requires proportionaly larger peak voltages across the line. The relatively large firing voltage of a spark gap, compared to other components used for transient protection applications, makes it possible to protect transmitting antennae with spark gaps.

In order to conveniently install the spark gap and minimize insertion losses, spark gaps are mounted in fixtures with coaxial connectors (types N, BNC, and UHF are particularly common). The coaxial spark gap was patented by Cushman (1960).

Coaxial spark gaps are available from:

    Cushcraft   (Blitz-bug^{TM})

    Genalex   (M O Valve, England)

    Lumex   (distributor for Cerberus of Switzerland)

    Reliance, division of Reliable Electric Comm/Tec

NEON LAMPS

Neon lamps have been used for many years as an indicator lamp in electronic circuits. The GaAs solid-state light emitting diode (LED) has made neon lamps obsolete for most electronic applications, although neon lamps continue to be commonly used for pilot lights that are connected to the mains. Neon lamps were also used as electronic circuit components, but this application is also obsolete owing to the superior performance of semiconductor light-emitting diodes. Neon lamps have been used for transient suppression by engineers who were familiar with them. The following advantages were claimed for neon lamps, compared to spark gaps:

1. Lower firing voltage (e.g. 60 volts DC, vs. 150 VDC, or more, for a spark gap)

2. Transparent case allows condition of neon lamp electrodes to be inspected

3. Neon lamps are available with a bayonet base that allows use of a socket, for convenient replacement.

4. Less expensive

There is no doubt that neon lamps do have a DC firing voltage that is smaller than that offered by nearly all spark gaps. However, C. G. Clare and Siemens each makes one spark gap (models CG-75, B1-C75 respectively) that have a 75 ± 15 volt DC firing voltage, and many manufacturers make models that have a 90 ± 20 volt DC firing voltage. During rapidly rising pulses, the maximum voltage across the neon lamp or spark gap will be several times the DC firing voltage. Therefore the advantage of the 60 to 70 volt DC firing voltage of neon lamps, is probably not appreciable when compared to low voltage spark

gaps (e.g. 75 to 150 volts DC).

The transparent glass case does allow the electrodes of a neon lamp to be inspected. Nearly all spark gaps (except Siemens Button Type gaps) have ceramic cases that are not transparent. One has to balance the advantage of being able to inspect the electrode inside a transparent case with the disadvantage that a glass case is probably easier to shatter during a transient than a ceramic case.

The use of neon lamps with bayonet bases in transient protection circuits for "ease of replacement" is not a good idea. The thin glass cases on neon indicator lamps can shatter during transients. Replacement of the lamp without the glass bulb is then difficult. Moreover, the inductance in the wires to the lamp socket is definitely worth avoiding.

The neon lamps are considerably less expensive than spark gaps (e.g. $0.30 for lamp, vs. $2.50 for spark gap), a feature that merits considerable attention. Although neon lamps are generally inferior to spark gaps for transient protection, the limited protection from a neon lamp may be preferable to no protection in an application where a spark gap is too expensive.

If neon lamps are to be used for transient protection, the model 5AH (formerly NE-83) appears to be desirable. It was designed for use in pulse circuits and with currents that are greater than other neon lamps. However, it is still more fragile than spark gaps.

DIELECTRIC STIMULATED ARCING

When certain crystalline dielectric materials are placed in an intense electric field, electrons are emitted from the surface of the dielectric. When this is used to initiate an arc in a gas that surrounds the dielectric, the phenomenon is called "dielectric stimulated arcing." Several investigators (Cooper and Allen, 1973; Brainard and Andrews, 1979) have constructed dielectric stimulated arcing spark gaps. The addition of the dielectric decreased the response time of the spark gap and reduced the spread of DC firing voltages that were obtained in a batch of spark gaps.

Kawiecki (1971) fastened thin strips of conductor that was connected to one electrode to the interior ceramic insulator wall of a spark gap. The conductor was parallel to the axis of the spark gap, and between 0.4 mm and 0.8 mm in width. He claimed that the enhanced electric field at the end of the strip was responsible for "stablizing the operating characteristics" and providing a much faster response. At a 10 kV/µs rate of rise a gap with a DC breakdown voltage of 250 volts will fire at 500 volts with the strips and 3000 volts without the strips. Dielectric stimulated arcing from the alumina ceramic wall might be responsible for the performance of these strips. Once the electrical discharge is established in the gas, the majority of the current is conducted in a discharge directly between the electrodes, and not involving the strips. This is due to the relatively low conductivity of the strips compared to the ionized gas.

Brainard, Andrews, and Anderson (1981) used pellets of zinc oxide varistor material between two spark gap electrodes. At low voltages the zinc

oxide varistor functioned in the usual non-linear manner. At higher voltages, an arc in air was initiated by an electrical discharge at the point contact between one electrode and a varistor pellet. The arc was claimed to protect the varistor from damage by excessive current.

VARISTORS

"Varistors" is the generic name for "voltage-variable-resistors." These devices obey the relation $V = I \times R$, where R is a function of V or I. Strictly speaking, all non-linear devices may be called "varistors." However, the conventional use of the word "varistor" is restricted to devices that dissipate energy in a bulk material (not a semiconductor junction), and for which R decreases as the magnitude of V (or I) increases. A varistor is inherently a bipolar device: its characteristics are not polarity dependent. Most varistors that are commercially available are fabricated from either (1) silicon carbide, (2) selenium, or (3) various mixtures of metal oxides (of which zinc oxide is the principal ingredient). Of these three kinds, the metal oxide varistor has the greatest non-linearity. The availability of zinc oxide varistors has essentially made the silicon carbide and selenium varistors obsolete. The metal oxide varistor is commonly denoted by the abbreviation "MOV." A typical (V,I) curve for a metal oxide varistor is shown in Fig. 4-1. The older silicon carbide and selenium varistors are commonly known in the U.S.A. by the General Electric tradenames "Thyrite" and "Thyrector", respectively.

At the present time, varistors are not precision components. The voltage, V_N, at which the current in the varistor is 1 mA, is used as a nominal conduction voltage. V_N is specified with a tolerance of $\pm$ 10% for MOV models with $V_N \geq 33$ volts. For smaller values of V_N, tolerances can be as large as $\pm$ 30%. Varistors do not have a sharp knee in their V vs. I plot, unlike silicon avalanche diodes, so the voltage across the varistor is not

Fig.4-1

+10A

300 VOLTS

−10A
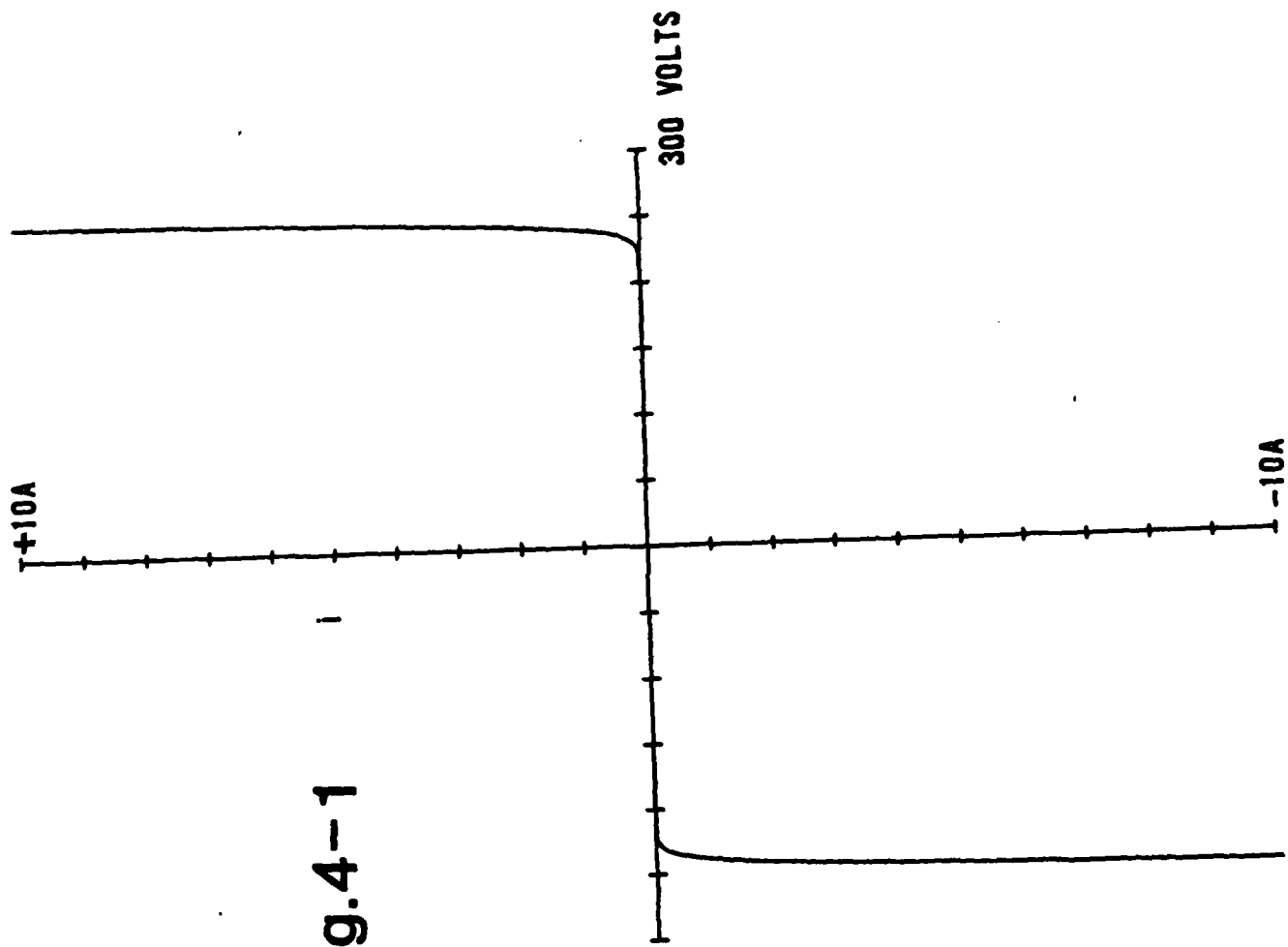
sufficiently independent of current to make varistors useful for voltage regulator applications.

Varistors degrade gradually when subjected to occasional pulse currents. The "end of life" is commonly specified when V_N has changed by ± 10 %. However, the varistor is still functional after the "end of life." In nearly all cases the value of V_N decreases with exposure to pulse currents. This aging effect manifests itself as an increase in leakage current at the maximum normal operating voltage in the system. The varistor is, like the gas tube, a consumable component. When subjected to excessive continuous currents, the varistor usually fails as a short-circuit. However, when "blasted" with excessive pulse current, the varistor may rupture and fail as an open-circuit.

Varistors that are designed for large pulse currents (more than 1 kA peak) are fabricated as a disk with a diameter of at least 14 mm. This large cross-sectional area makes the current per area small. However, this also makes the capacitance of the varistor rather large. Varistors have the largest capacitance of the common non-linear transient protection devices; values of 1 to 10 nF are common. This capacitance is not always a bad feature. For example, the shunt capacitance of varistors can be used to construct non-linear low-pass filters (Campi, 1977).

One will find remarks in the literature about the "slow" response of MOVs. Often response times of the order of a few nanoseconds to tens of nanoseconds are given as typical of MOVs. The actual response time of the varistor is less than 0.5 ns (Philip and Levinson, 1981); the apparent response time is due to parasitic inductance in the package and leads.

There is general agreement that varistors are the best of the presently-
available non-linear devices for protection of AC power lines. They are also
suitable for suppression of transients caused by switching inductive loads
off. Other applications are, of course, also possible.

When the (V,I) characteristic curve is plotted on a log-log scale, Fig.
4-2, three characteristic regions of operation become apparent (Harnden, et
al., 1972). At very small currents, less than 10^{-4} amperes, the varistor
behaves as a simple resistor, called R_{leak}. At very large currents, more than
10^2 amperes, the varistor response is dominated by the bulk resistance of the
device, R_{bulk}. In between, the varistor obeys the following ideal relation.

$$I = k V^{\alpha} = (V/D)^{\alpha}$$

The circuit diagram that corresponds to this model of a varistor is shown in
Fig. 4-3. The inductance shown in Fig. 4-3 is due to packaging and the length
of the leads in a particular application.

The value of α characterizes the non-linear (V,I) characteristic. An
ordinary resistor would have $\alpha = 1$. As a general rule, the larger the value
of α, the "better" the varistor. Modern metal oxide varistors have values of
α of at least 25. The varistor relation, given above, with the parameter k
often causes problems when implemented in a computer program. The value of k
is often less than 10^{-100}, which causes "underflow." This difficulty is
eliminated by using the form of the varistor relation, given above, with the
parameter D.

The value of α is determined with the following relation.
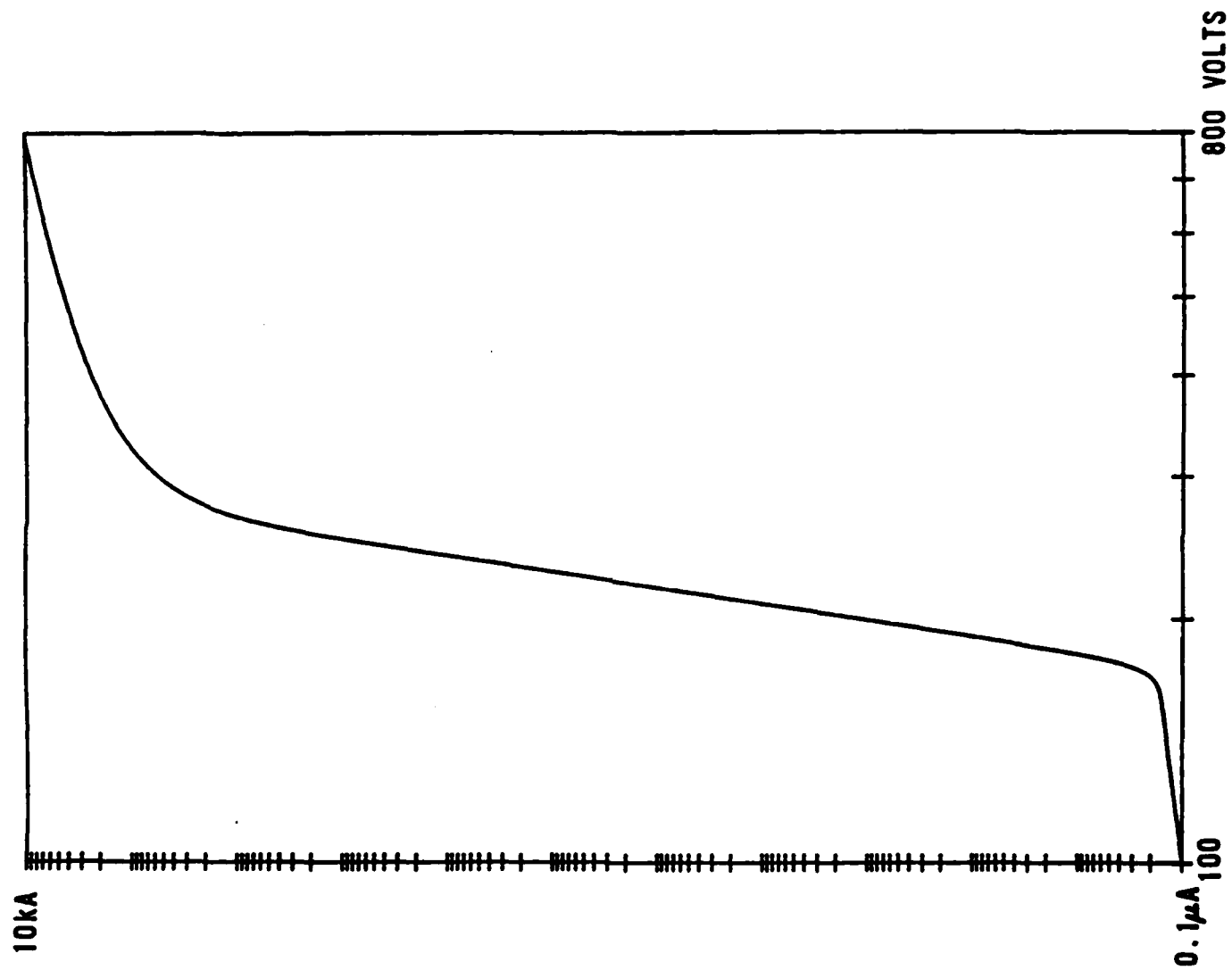
Fig.4-2

Fig.4-3

VARISTOR EQUIVALENT CIRCUIT MODEL
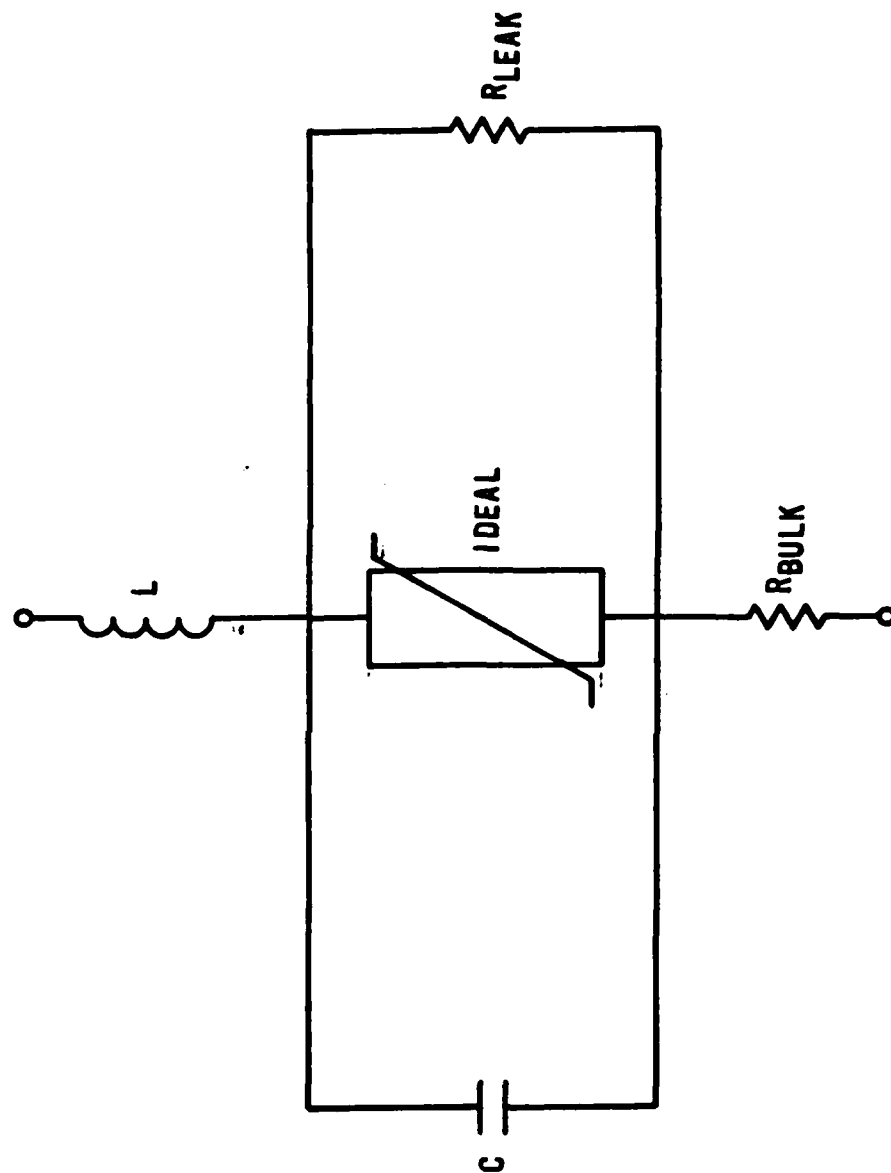
$$\alpha = \frac{\log(I_1/I_2)}{\log(V_1/V_2)}$$

If either (V_1, I_1) or (V_2, I_2) is in the region where leakage resistance or bulk resistance has a significant effect, then the value of α calculated from the above relation will be too small. Early General Electric varistor data sheets (e.g. Type ZA, May 1975; Type MA, 1976) specified the value of α for I at 0.1 mA and 1 mA. This appears to be a good way to ensure that everyone agrees on how to determine the value of α. This is important if the value of α appears in a specification, since errors of ± 1 can easily be made in determinations of α, depending on technique. To prevent quibbles between vendor and user on marginally acceptable (or marginally unacceptable) varistors, it is prudent to define the method for determinations of α in the specifications.

In addition to the effects mentioned above, there is a temperature dependence of the (V,I) characteristics. The varistor voltage at a constant current (I > 1 mA), changes by about -0.04% per kelvin increase in temperature (Smith and McCormick, 1982, p.50). This is of little consequence in transient protection applications, but if the overstress is sustained, this property can produce thermal runaway.

The value of V_N, the voltage between the varistor terminals at a 1 mA constant current, and the value of α are two parameters that completely specify an ideal varistor. The values of C, R_{leak}, and R_{bulk} complete the model. Most manufacturers, however, specify V_N, a typical value of C, and the maximum value of V at some arbitrarily large pulse current.

In applications where exposure to large pulse currents are anticipated or where reliability is a principal concern, there are two particular points to consider. First, one should select a model of varistor with the largest diameter that will fit in the available space. For a given current, the current density is less in varistors with larger diameters. This allows the varistor to absorb larger transients without significant degradation. Second, one should specify the minimum initial value of V_N to be at least a factor of 1.25 above the peak voltage expected during normal operation of the system. This gives a larger margin for degradation before the varistor will interfere with normal system operation. If this value of V_N is too large to completely protect the system, another varistor with a smaller value of V_N can be connected downstream (toward the system to be protected) with a suitable impedance (inductor or resistor) in series between the two varistors.

Another possibility for increasing reliability during and after exposure to large pulse currents is to connect two varistors in parallel. This is not recommended, because non-linear devices do not share current evenly. As an example, consider two varistors in parallel, with the following properties.

V_1: V_N = 78 volts α = 32

V_2: V_N = 86 volts α = 28

These two varistors both meet specifications for V_N = 82 volts ± 5 %. When 115 volts is placed across these two varistors the current in V_1 will be about 250 amperes, while the current in V_2 will be about 3.4 amperes. In this example 98.6% of the total current flows in V_1. The second varistor is ineffective in this combination, however ±5% is good matching for production varistors. The first varistor has an initial value of V_N that is smaller than that of the second varistor. Since the first varistor takes most of the

stress, it will be degraded more than the second varistor. Therefore, with increasing age, the first varistor's value of V_N will decrease faster than the second varistor's value of V_N.

The ineffectiveness of placing two nearly identical varistors in parallel is not a problem that is unique to varistors. The same effect occurs with other highly non-linear devices such as zener and avalanche diodes.

The above example of two varistors in parallel is not quite realistic in that the bulk resistance of the varistor was ignored. When large currents flow in V_1 the bulk resistance will increase the voltage across its terminals. This "extra" voltage will increase the current in V_2 (Smith and McCormick, 1982, p.68).

Metal-oxide (ZnO) varistors were invented in Japan by Matsushita and licensed to other companies. The fabrication of ZnO varistors has been described by Matsuoka, et al. (1970). The following companies, in alphabetical order, sell metal-oxide varistors in the U.S.A.

CKE, Inc.

Collmer Semiconductor (distributor for Fuji Semiconductor)

General Electric

Marcon America Corp.

Mepco/Electra (N. V. Philips subsidiary)

Panasonic (Matsushita Electric Corp. of America)

Siemens

Thompson-CSF

Teledyne Relays

Victory Engineering Corp.

The following companies sell silicon carbide varistors in the U.S.A.

Carborundum, Electric Products Div.

Midwest Components Inc.

AVALANCHE AND ZENER DIODES

Silicon diodes that are intended for use in the reverse-breakdown region
are familiar components to analog electronic circuit designers. While these
components are commonly called "zener diodes," only devices with a reverse
breakdown voltage of less than about 5 volts usually use the Zener mechanism.
Diodes with a reverse breakdown voltage of more than 8 volts use the
avalanche mechanism. For diodes with a reverse breakdown value between about
5 and 8 volts, both mechanisms operate. One can quickly identify the
mechanism of a particular diode in the laboratory with the following rule.
In Zener breakdown the magnitude of the conduction voltage decreases as the
temperature increases; for avalanche breakdown, the magnitude of the
conduction voltage increases as temperature increases. Both avalanche and
zener diodes are widely used as voltage regulators because in the reverse
breakdown region, the magnitude of dV/dI is very small. A (V,I) curve for a
typical avalanche diode is shown in Fig. 5-1.

The relationship between voltage and current for zener and avalanche
diodes can be fit to the same mathematical expression that was used for
varistors:

$$I = (V/D)^{\alpha}$$

Fig.5-1
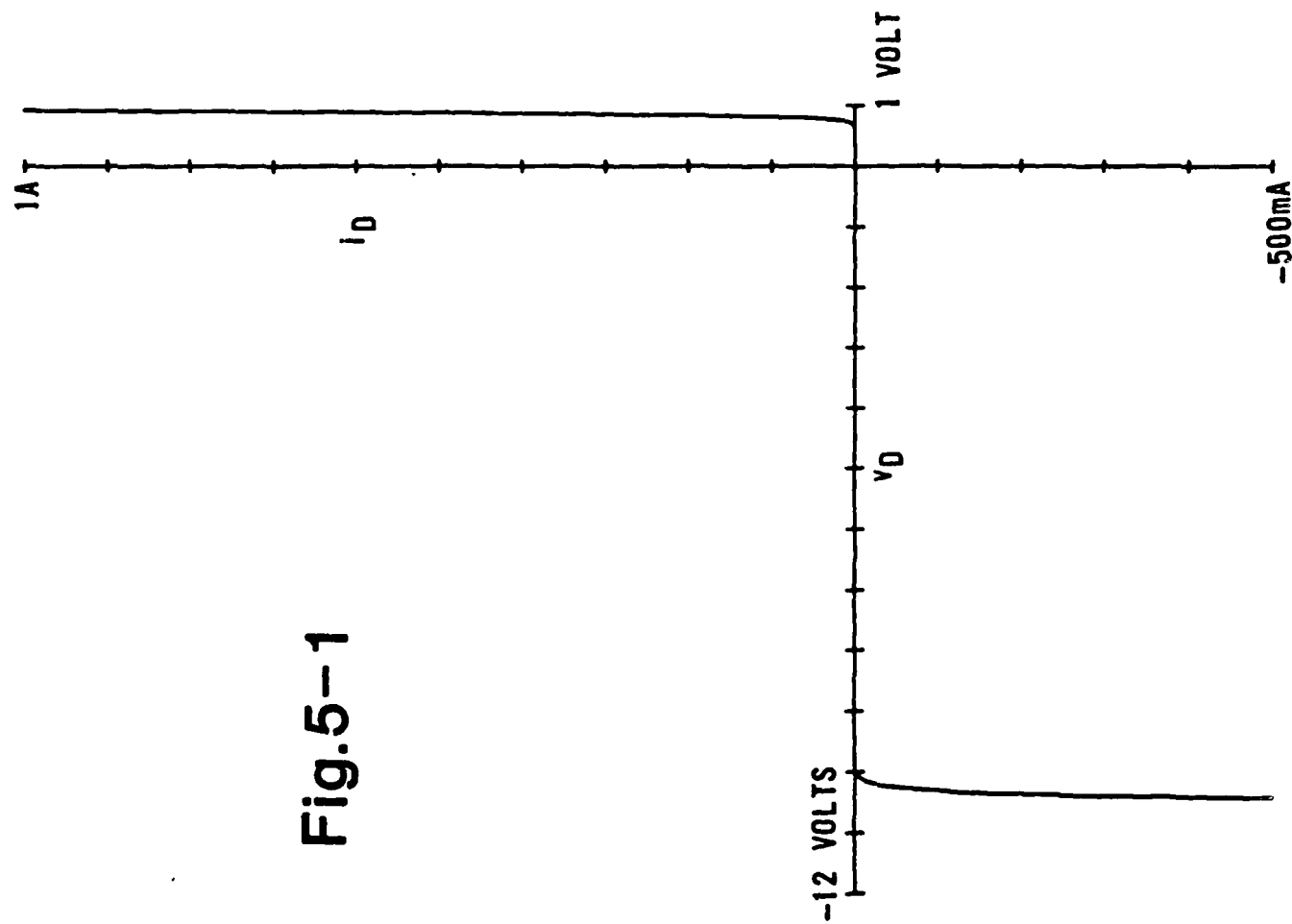
The value of α for some particular zener and avalanche diodes is given in the following table.

Part number	Nominal Parameters			α	type
1N3821	3.3 V	1	W	7.7	zener
1N3822	3.6 V	1	W	8.4	zener
1N5334	3.6 V	5	W	7.6	zener
1N3993	3.9 V	10	W	8.4	zener
1N5524	5.6 V	0.5	W	209	avalanche
1N3827	5.6 V	1	W	448	
1N5339	5.6 V	5	W	467	
1N4736	6.8 V	1	W	398	
1N5063	6.8 V	3	W	154	
1N4954	6.8 V	5	W	144	
1N6267*	6.8 V	5	W	130	
1.5KE11C	11 V	5	W	55	avalanche
1N5348	11 V	5	W	126	avalanche
1.5KE33C	33 V	5	W	214	avalanche
1N5364	33 V	5	W	470	avalanche
1N6288*	51 V	5	W	513	avalanche
1N4989	200 V	5	W	394	avalanche
1N6303*	200 V	5	W	662	avalanche

* member of 1.5KE series of transient suppression diode

From inspection of these data it is evident that zener diodes have much smaller values of α than avalanche diodes. This will be important in protecting integrated circuits with small operating voltages, such as high speed CMOS. Avalanche diodes have too large a breakdown voltage to be

suitable for protection of all low voltage integrated circuits. Zener diodes have the appropriate range of breakdown voltages, but are not highly non-linear.

Zener and avalanche diodes are quite polarity sensitive. In the forward-biased direction they act like an ordinary silicon rectifier, when the reverse-biased they have the characteristic shown in Fig. 5-1. To clamp overvoltages of either polarity, two zener or avalanche diodes are connected back to back in series, as shown in Fig. 5-2. It is common for the two diodes in Fig. 5-2 to have the same breakdown voltage (within ±10% tolerances), but this is not required in order to use the circuit of Fig. 5-2. Because this configuration is so common in transient protection applications, many manufacturers provide two identical avalanche diodes connected back to back inside one package. Such devices are called "bipolar" avalanche diodes. This packaging concept has the advantage of reducing parasitic inductance in series with the diodes, as well as minimizing both the assembly time and mass of the circuit.

Avalanche diodes are particularly fast-responding. One well-known manufacturer of transient suppresor diodes claims a response time of less than one picosecond for his product. This is rubbish, because a signal traveling at the speed of light in vacuuo would require about 30 picoseconds to traverse the length of the plastic package that contains the diode. In practice, the response time is determined by the parasitic inductance of the package and leads. This parasitic inductance can produce appreciable overshoot for 50 ps after the beginning of a rectangular pulse (Clark and Winters, 1973).
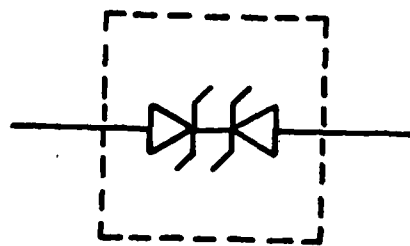
Fig.5-2

Diodes that are able to tolerate large pulse currents also have a large capacitance, as a consequence of the large cross-sectional area that is necessary to reduce the current density. It is not uncommon to find transient suppression diodes with capacitance values between 500 and 1000 pF. Such values make these diodes alone unsuitable for use in circuits with high-frequency signals. However, these diodes can be placed in series with forward-biased switching diodes, which have a capacitance of the order of 50 pF, to reduce the total shunt capacitance, as shown in Fig. 5-3a. Forward-biased diodes with large pulse current ratings are relatively slow to conduct, a phenomemon that will be discussed in Chapter 6. While the circuits of Fig. 5-3 do decrease the shunt capacitance, they introduce another problem: that of slower response (Clark and Winters, 1973, p.83).

The use of switching diodes in series with avalanche diodes prevents current flow when the avalanche diode is forward-biased. Therefore, the bidirectional clamping circuit of Fig. 5-2 will not work when switching diodes are included. To clamp overvoltages of either polarity, use one of the circuits that are shown in Fig. 5-3 (Popp, 1968, p.27; Clark, 1975). The circuit of Fig. 5-3b is less expensive than the circuit of Fig. 5-3a since the former uses only one avalanche diode, which are much more expensive than rectifiers or switching diodes. However, when the circuit of Fig. 5-3b is fabricated from discrete components, the additional parasitic inductance of the extra switching diode in the conducting path may be a serious disadvantage.

(a)

(b)
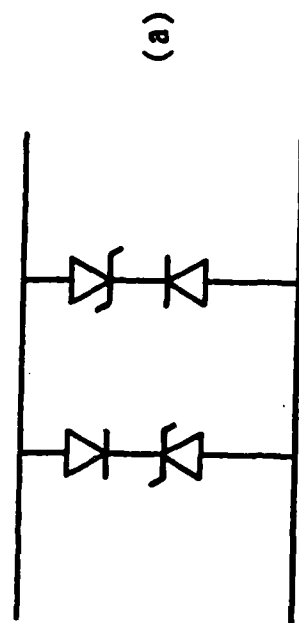
Fig.5-3
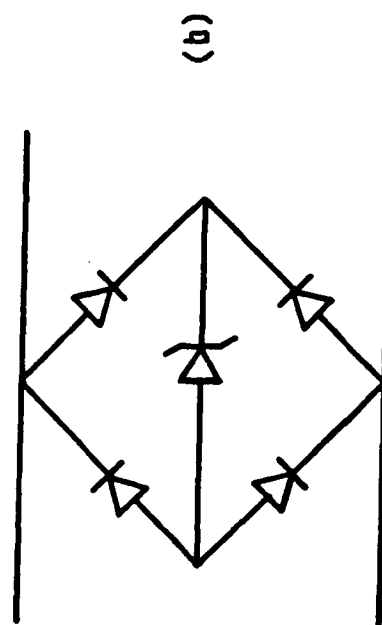
Steady-state power dissipation ratings of common zener and avalanche diodes are between 0.5 and 5 watts. During brief transients (durations of a few microseconds or less), most diodes can have an instantaneous power level that is at least a hundred times greater than their steady-state rating. Even so, the peak current of most small diodes is limited to a few tens of amperes. This limited pulse current is one of the major shortcomings of diodes as non-linear transient suppression devices. Of all of the common TPDs, zener and avalanche diodes have the smallest energy absorption capability.

Silicon diodes do not exhibit degradation of electrical parameters after prolonged service (unlike gas tubes or varistors). This fact is often stated as a reason to use silicon diodes rather than varistors. However, one should also consider the fact that the maximum tolerable surge current for a zener or avalanche diode in reverse breakdown is rather small when compared to tolerable surge currents for varistors and spark gaps.

Diodes usually fail as a short-circuit. However, under sustained large currents the device can rupture an epoxy package that are used for commercial grade devices. If a short-circuit failure mode is essential, one should use devices in a hermetically sealed metal package (e.g. DO-13). In some tests with large steady-state currents avalanche diodes were observed to fail with a behavior similar to that of second breakdown. In one such test 300 mA was passed through a 1N5364 diode. The power dissipation was 11 watts, substantially greater than the 5 watt maximum steady-state rating. After about 22 seconds at 300 mA, the diode failed. Thereafter the former 33 volt diode had a characteristic of a constant voltage of 15 volts at currents

greater than about 10 mA, and of a 2 kΩ resistor at smaller reverse currents.

A particular disadvantage of avalanche diodes is that, when operated at relatively small reverse currents (e.g. 1 μA to 1 mA), many avalanche diodes produce significant noise with a bandwidth from less than 1 Hz to over 0.5 MHz. In fact, special avalanche diodes are used as a common source of "white noise". Care must be taken to insure that the normal operating voltages of the system do not push the avalanche diode into a weakly conducting state.

Special avalanche diodes with breakdown voltages less than 6 volts are available, and are widely used as voltage reference circuits in battery operated equipment where small diode currents and a small value of dV/dI are required. These low voltage avalanche devices have relatively small power ratings. However these devices may be attractive for protecting high-speed CMOS and other low-voltage integrated circuits.

Special transient suppressor diodes are available from several different manufacturers. How do these special transient suppression avalanche diodes differ from "ordinary" avalanche diodes with large steady-state power ratings? It appears that the special diodes have larger heat sinks, which are often made of unusual materials, than ordinary diodes. Several different diodes were examined with x-ray fluorescence to determine the heat sink material.

Manufacturer and Part Number		Heat Sink Material
Special Transient Diodes		
General Semiconductor Industries 1.5KE33C		Ag
General Semiconductor Industries P6KE18C		Ag
Motorola	1.5KE18	Ni plated Cu
Semicon	1.5SE18C	Al
Ordinary Diodes		
Fairchild 1N4004 rectifier		Cu, Fe, Ni
Unitrode UZ5833 miniature avalanche diode		Rh plated Mo

Silver and copper have the greatest and second-greatest thermal conductivity of any known material, and would be the best choices for rapidly removing heat from the junction. However, the coefficient of thermal expansion of neither silver nor copper are well matched to silicon. Reynolds (1972) used a tungsten heat sink in a transient supressor avalanche diode. Most power semiconductors use a molybdenum internal heat sink.

Special avalanche diodes that are designed and specified for transient suppression are available from many different companies. These devices are available by specifying either JEDEC registered part numbers or proprietary part numbers. The following list gives three commonly used JEDEC series of avalanche transient suppression diodes.

1N5629 to 1N5665	6.8 to 200 volts	DO-13 hermetic package
1N6036 to 1N6072	7.5 to 220 volts BIPOLAR	DO-13 hermetic package
1N6267 to 1N6303	6.8 to 200 volts	epoxy package

Avalanche transient suppression diodes are available from the following companies (which are listed alphabetically) in the U.S.A.

General Semiconductor Industries

Motorola

Semicon

Siemens

Thompson-CSF

Unitrode

Avalanche transient suppressor diodes in series with forward-biased diodes in a single package are available from the following companies:

General Semiconductor Industries LC and LCE series

Semicon SLC and SLCE series

Thompson-CSF PFC series

Avalanche diodes with breakdown voltages between 3.3 volts and 10 volts are available from several sources:

1N5518 - 1N5530 series	Motorola, KSW Electronics
1N6082 - 1N6091 series	KSW Electronics, TRW Semiconductor
LVA series	TRW Semiconductor

DIODES

Silicon diodes and rectifiers are also useful non-linear devices for transient protection. These devices are are usually employed so that they conduct when forward-biased, unlike zener or avalanche diodes. A (V,I) characteristic curve for a typical diode under steady-state conditions is shown in Fig. 6-1.

The relation between current, i_D, in a semiconductor diode and potential difference, v_D, across a semiconductor diode is usually given in the form of Eqn. 1.

$$i_D = I_S \left(\exp(v_D/mV_T) - 1 \right) \tag{1}$$

I_S, V_T, and m are three parameters that characterize the diode's (V,I) relationship.

A common value of I_S is on the order of 1 nA. With a knowledge of semiconductor physics one can derive Eqn. 1 and the relation

$$V_T = kT/e$$

where k is Boltzmann's constant (about 1.38×10^{-23} joule per kelvin), e is the elementary charge (about 1.60×10^{-19} coulombs), and T is the temperature of the diode (in kelvin). The value of kT/e is about 1/40 volt or 25 mV at room temperature, about 300 kelvin. An empirical constant, m, which is usually between 1.1 and 2.0, is multiplied times V_T to give better agreement between the current predicted from Eqn. 1 and experimental data.
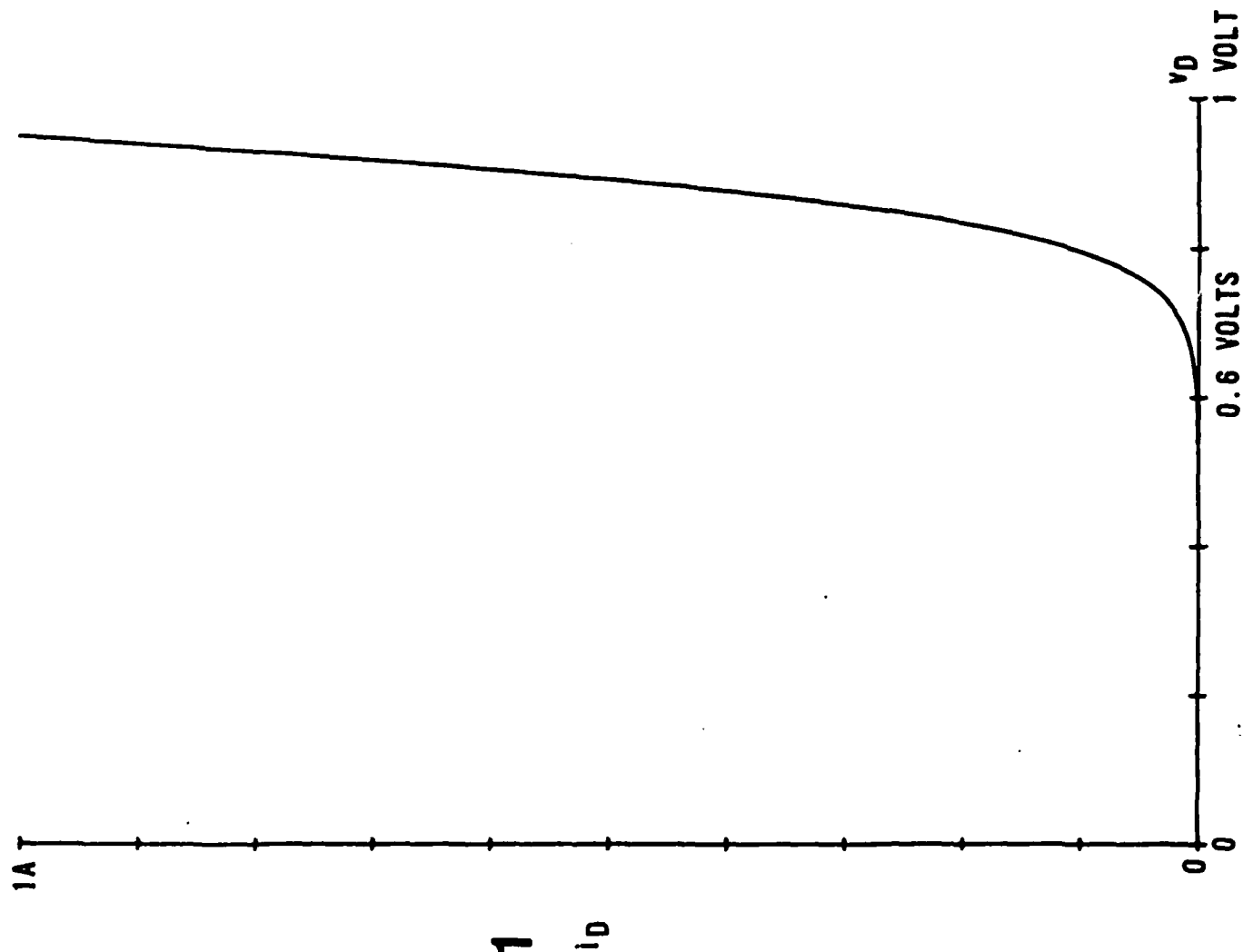
Fig.6-1

The following table contains values of I_S and mV_T for some representative diodes at room temperature.

Diode Model Number and type	I_S	$m V_T$ at room temperature
1N4447 switching diode	2.7 nA	47 mV
1N4004 rectifier 1 A	5.0 nA	47 mV
1N5624 rectifier 3 A	9.5 nA	45 mV
1N3822 zener 3.6 V	0.2 pA	28 mV
1N5348 avalanche 11 V	0.1 pA	29 mV
1N5364 avalanche 33 V	0.1 pA	29 mV

When one examines a plot of $(v_D, \log_{10}(i_D))$ for a real diode, Fig. 6-2. (We emphasize that Fig. 6-2 is not a plot of Eqn. 1.) One finds that the curve "rolls over" somewhere in the vicinity of 0.8 to 0.9 volts. The different slope at greater values of v_D is due to two different causes: temperature and bulk resistance. As the current becomes large, appreciable power ($P = v_D i_D$) is dissipated inside the diode. This causes the temperature of the diode to increase, and thus V_T will increase. At a constant current the voltage across the diode junction v_D is proportional to V_T. Thus, at a given current, hotter diodes have larger values of v_D. The resistance of the bulk semiconductor material produces a voltage drop across the diode's terminals that is in addition to the voltage v_D given in Eqn. 1. Remarkably, this behavior at large currents can be modeled by using Eqn. 1 with I_S and V_T treated as constants plus an additional voltage drop due to a resistance R, as shown in Eqn. 2. Typical values of R are 0.05 ohm for a rectifier with a maximum
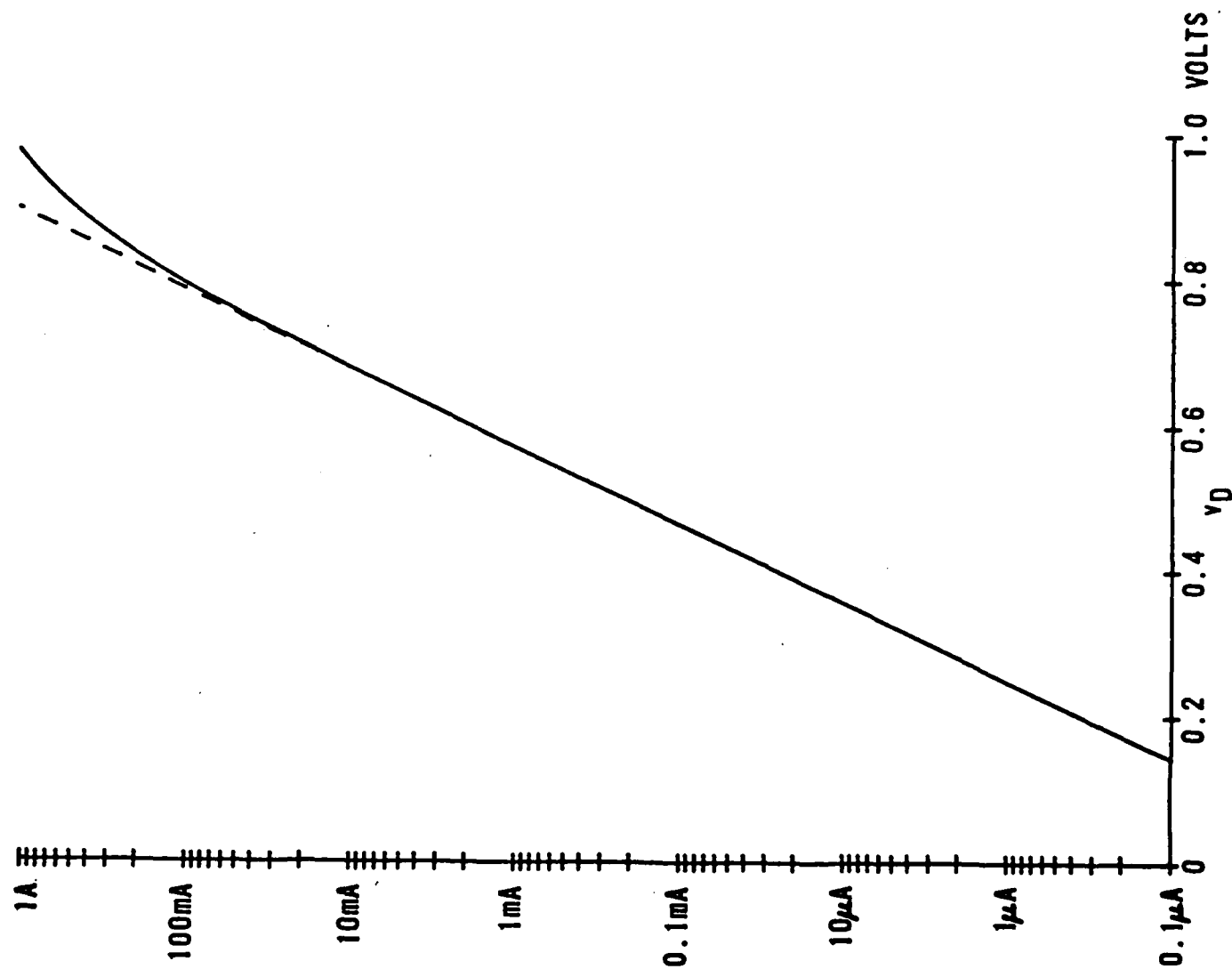
Fig.6-2

current of 1 ampere.

$$v_D = m \, V_T \, \ln(i_D/I_S) \; + \; i_D \, R \tag{2}$$

One of the advantages of forward-biased diodes is that they clamp transient currents at a very small voltage, about 0.6 to 2 volts. This is the smallest clamping voltage of any presently available semiconductor. Diode clamps are particularly useful to prevent the base-emitter junction of a transistor from being reverse-biased.

Diodes do not switch instantly. When a semiconductor diode is suddenly switched from zero current to a strongly forward-biased condition (e.g. 500 mA constant current), the voltage across the diode reaches a peak value that can be more than 10 times the steady-state value. The time required for this transient voltage to decay is known as the "forward-recovery time." This phenomenon has been discussed by Cooper (1962). An "ordinary" 1N4007 industry-standard rectifier was connected across the output terminals of a Hewlett-Packard 214B pulse generator so that the diode was forward-biased. The voltage across the rectifier was monitored on a Tektronix 7844 dual-beam oscilloscope. The open-circuit voltage (without the rectifier) was a 60 volt rectangular pulse that was on for 2 µs and off for 500 µs. Peak current in the diode was 1.4 A. A plot of the voltage vs. time is shown in Fig. 6-3. In this situation the voltage across the forward-biased diode was greater than 2 volts for 0.4 µs. This experiment shows that forward-biased diodes do not always have a 0.6 or 0.7 volt drop across them, as is assumed in undergraduate electronics classes for engineering students. The forward-recovery time is
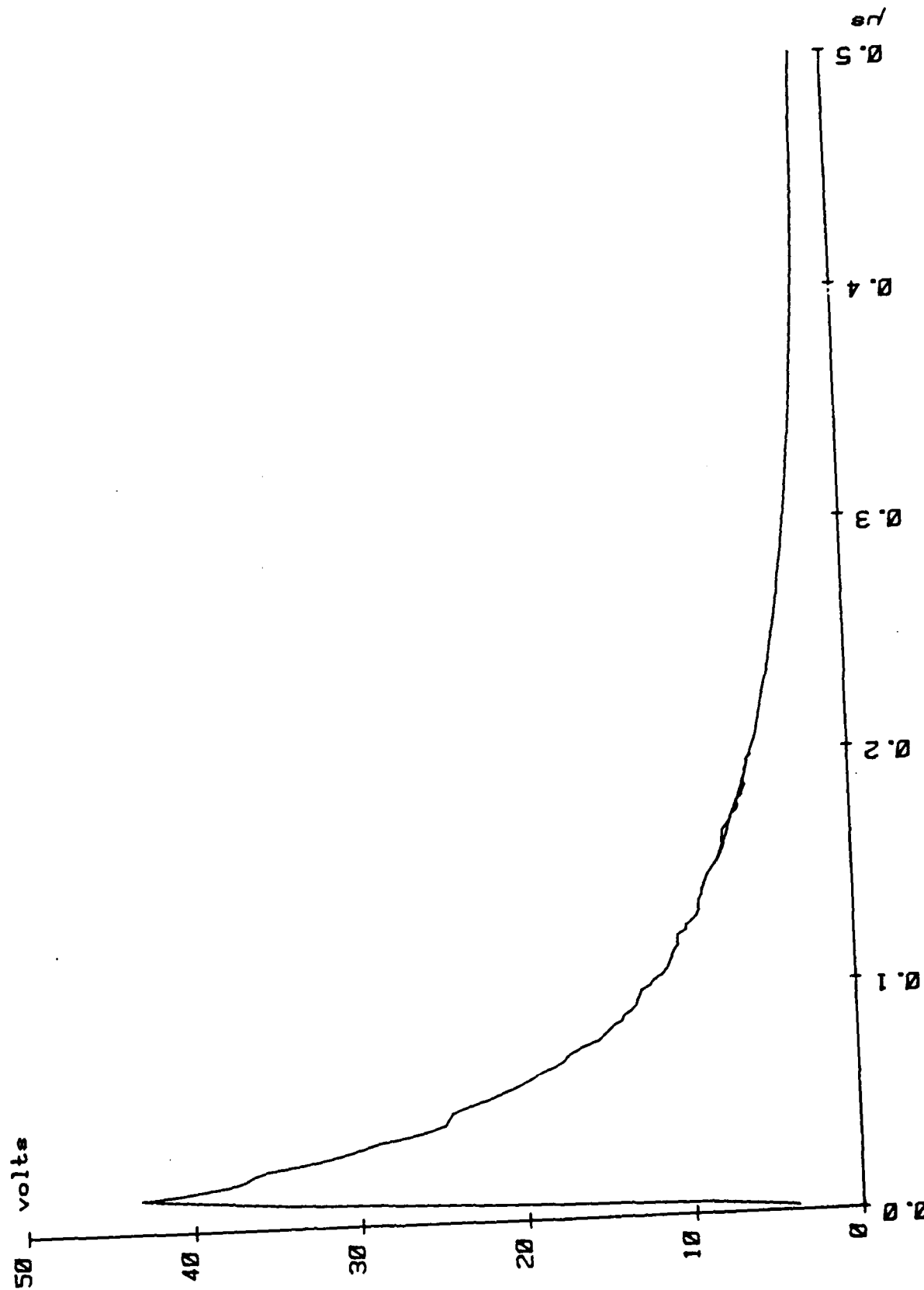
Fig.6-3

particularly long in high-voltage rectifiers (e.g., the 1N4007 has a reverse
breakdown voltage of at least 1000 volts). Therefore, when diodes with a fast
response are needed in transient protection circuits, the diodes will have a
small reverse-breakdown voltage specification (e.g. 75 to 200 volts).

When a conducting diode is suddenly reverse-biased, the current does not
drop to zero instantly. For a brief period (which can be as long as tens of
microseconds, depending on the diode construction), the diode will conduct
current in the reverse direction. The duration of this brief period during
which the diode fails to rectify is known as the "reverse-recovery time." In
most cases the reverse-recovery time is greater than the forward-recovery
time, so the reverse-recovery time can be used as a measure of the diode's
speed.

We shall discuss four sub-classes of silicon diodes in this section.
1. core switching diodes (e.g. 1N4447)
2. ordinary rectifiers   (e.g. 1N4004)
3. fast-recovery rectifiers
4. Schottky barrier diodes

The 1N914 and 1N916 switching diodes were developed for use in computer
core memories. The 1N4447 is the same semiconductor as the 1N916A, but the
1N4447 has an improved internal heat sink so that it can tolerate about twice
the power or current of the 1N916A. The 1N4447 diode features:

a small capacitance (less than 2 pF at 0 volts across the diode),

fast response (reverse recovery time < 4 ns for $I_F = I_R = 10$ mA),

quite inexpensive (about $0.15 each in quantities of 100),

relatively low leakage devices (less than 25 nA at 20 V reverse bias). The JEDEC ratings specify a maximum DC current of 0.4 A, 1 A for 1 s, and 4 A for 1 µs. (We mention in passing that it is unusual to find 1 µs surge current ratings in JEDEC specifications. The reason is that in the original application for these diodes, which was switching magnetic core memories, the current was applied in brief pulses instead of a continuous current.)

The 1N4448 is an improved version of the 1N914B. The 1N4448 is similar to the the 1N4447, but the 1N4448 has less bulk resistance and slightly greater capacitance (less than 4 pF at 0 V across the diode).

The 1N4148 is an industry-standard switching diode and may be somewhat easier to obtain than the 1N4447 or 1N4448 types.

A rectifier is a diode that is designed to conduct steady-state currents of the order of 1 A or more. Ordinary rectifiers are inexpensive and have large surge current ratings (e.g. a rectifier that is rated to carry 1 A steady-state can tolerate a peak current of 30 A for one pulse of half-wave rectified 60 Hz sinusoidal waveform), but can be slow (e.g. reverse recovery times of a few tens of microseconds). The typical capacitance of a 1N4004 rectifier at 0 V is about 15 pF.

Fast-recovery rectifiers overcome some of the disadvantages of ordinary rectifiers, but are more expensive and have a larger capacitance and larger reverse leakage current than ordinary rectifiers of the same steady-state current and reverse breakdown ratings. State of the art fast-recovery rectifiers have reverse recovery times of the order of 30 ns. Other "fast-recovery" rectifiers have reverse recovery times of the order of 200 ns.

Schottky barrier diodes and rectifiers are among the fastest diodes available. However they have a rather small reverse breakdown voltage rating (20 to 80 volts is a common range) and a rather large capacitance (200 pF at 0 V is typical for a diode that is rated to conduct 1 A steady-state). Schottky diodes begin to conduct at about 0.3 volts forward-bias, compared to about 0.6 volts for regular silicon diodes. Schottky diodes also have a much larger leakage current than regular silicon diodes when reverse-biased.

Two diodes connected in anti-parallel form a bipolar clamping circuit, as shown in Fig. 6-4a. Such a configuration is available in a single package from several companies (e.g. General Semiconductor Industries GSV 103, General Instrument SV100-1, Schauer SV1). Such devices are often called "silicon varistors", but the use of the word "varistor" to describe a semiconductor junction should be deprecated.

When a larger clamping voltage is desired, but the large capacitance of zener or avalanche diodes is unacceptable, one can form series strings of two or more forward-biased diodes, as shown in Fig. 6-4b. For bipolar protection, two series strings are connected in anti-parallel. Such a configuration is available in a single package from several companies (General Semiconductor

(a)
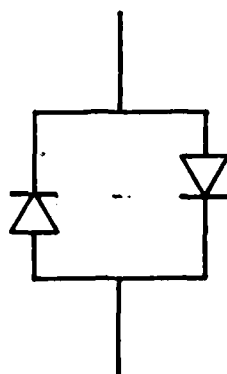
Fig.6-4

(b)

Industries GHV-2 to GHV-16; Semicon LCC-2 to LCC-16).

LOW-LEAKAGE DEVICES

In many situations in electronic instrumentation a very large input impedance (e.g. 10^9 Ω to 10^{12} Ω) is required. An operational amplifier with a junction field-effect transistor (JFET) or metal-oxide semiconductor field-effect transistor (MOSFET) input stage is commonly used. Transient protection of the input terminals is constrained by the specification that leakage currents during normal opertion must be of the order of 10^{-12} ampere or less. Very few devices can meet this requirement.

Spark gaps can meet this leakage current requirement, but the firing and glow voltage of a spark gap will kill a semiconductor operational amplifier. Further, spark gaps can be slow to conduct. Thus spark gaps alone are inadequate protection.

Among semiconductor devices, it has been customary to use the gate to channel diode of a JFET as a low-leakage diode, as shown in Fig. 6-5. These diodes are used in the forward-bias mode during transient protection. Two JFETs connected in anti-parallel as a bipolar diode are available in a single package from several manufacturers (Siliconix DPAD1, Intersil ID100). For unipolar operation, one can connect a JFET that has a gate to channel leakage current that meets the following specification:

$I_G < 1$ pA when $V_{GS} = 0$ at a temperature of 25 celsius.

Fig.6-5

N-CHANNEL
JFET CONNECTED
AS A DIODE
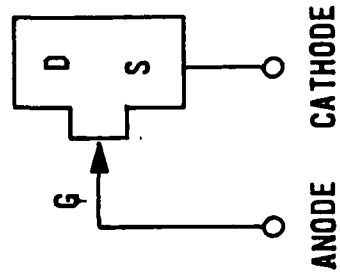
A popular model that meets this specification is the 2N4117A or FN4117A (available in a TO-72 metal can package). The same device is available in a less expensive TO-92 plastic package as part number PN4117A. One can also use JFETs that have been internally connected as a diode (Siliconix PAD1, JPAD5; Fairchild FJT1100).

The maximum steady-state gate current for the 2N4117A in a TO-72 metal can package is 50 mA. This is a relatively large gate current, compared to most other JFETs. The (V,I) characteristic curve for the 2N4117 gate to channel diode is described by Eqn. 2 with the following typical values of the parameters:

$$I_S = 0.0019 \text{ pA} \qquad V_T = 28 \text{ mV} \qquad R_{bulk} = 20 \text{ } \Omega$$

In an important article, Damljanovic and Arandjelovic (1981) suggested that GaAsP diodes be used for protection of input terminals of amplifiers with large input impedance. They found that at 0.44 volts forward bias, a red GaAsP light-emitting diode (LED) had over seven orders of magnitude less current than a base-emitter junction of a Si NPN transistor. A green LED was even better. This difference is due to both lesser values of reverse saturation current and greater threshold voltage for conduction in GaAs compared with Si.

In order to confirm and extend this work, four parts were chosen to test:

1. Optron OP290, an infrared LED that is designed for operation with large current pulses (maximum 5 A for 250 μs on, 20 ms off) and 125 mA maximum steady-state.

2. General Instrument MV5053, a red LED that has a relatively large steady-state current rating (100 mA max) and is quite inexpensive (about $0.25 each in small quantities).

3. General Instrument MV5094, a bipolar red LED. This device has a maximum steady-state current rating of 70 mA, and is relatively expensive (about $1.25 each in small quantities). This device may be preferable to the MV5053 only when the available space is small, since just one MV5094 is required for bipolar protection.

4. General Instrument MV5253, a green LED that is specified for 35 mA maximum steady-state current. Green LEDs with greater current ratings are not commercially available.

All three of the visible LEDs listed above are rated for a 1 A repetitive pulse current of 1 μs on, 333 μs off.

The relationship between current and voltage in these LEDs and in a JFET gate to channel diode are described by Eqn. 2 with the following typical values of parameters:

Model and type	I_S	m V_T at room temperature	V @ I =10 mA
PN4117A JFET	1.9×10^{-15} A	28.3 mV	1.03 volts
OP290 infrared	3.4×10^{-13} A	51.5 mV	1.24 volts
MV5053 red	1.2×10^{-17} A	47.4 mV	1.63 volts
MV5053 green	1.1×10^{-19} A	48.8 mV	1.90 volts

Fig. 6-6 shows the (V,I) characteristic curve for the LEDs and a PN4117 JFET
used as a diode. The relatively large bulk resistance of the JFET makes it
inappropriate for use where the current could exceed about 100 mA. The red
and green LEDs have a much smaller current at a given voltage than Si diodes,
and are suitable for protecting electrometer inputs, as suggested by
Damljanovic and Arandjelovic (1981).

PIN DIODES

A PIN diode has a layer of intrinisic (I) semiconductor material between
the positive (P) and negative (N) regions. Such diodes are usually employed
for switching very high frequency signals ($\nu > 200$ MHz) or for photo-sensitive
applications. We are interested in PIN diodes because they have extremely
small capacitance (0.2 to 2.0 pF) and, therefore, might be useful to protect
input terminals of RF receivers. Most other semiconductor devices are
unsuitable for such applications owing to an unacceptably large capacitance.

Singletary, Clive, and Hasdal (1971,p.31) suggested using a PIN diode in
series with a zener diode to reduce the effective capacitance. This valuable
suggestion appears to have been ignored for a decade. Lesinski (1983)
patented a circuit that used a 1N5719 PIN diode in series with a biased
avalanche device to provide transient protection for a UHF signal conductor.
Apparently PIN diodes have not been characterized with large pulse currents
that would be encountered during transient suppression.
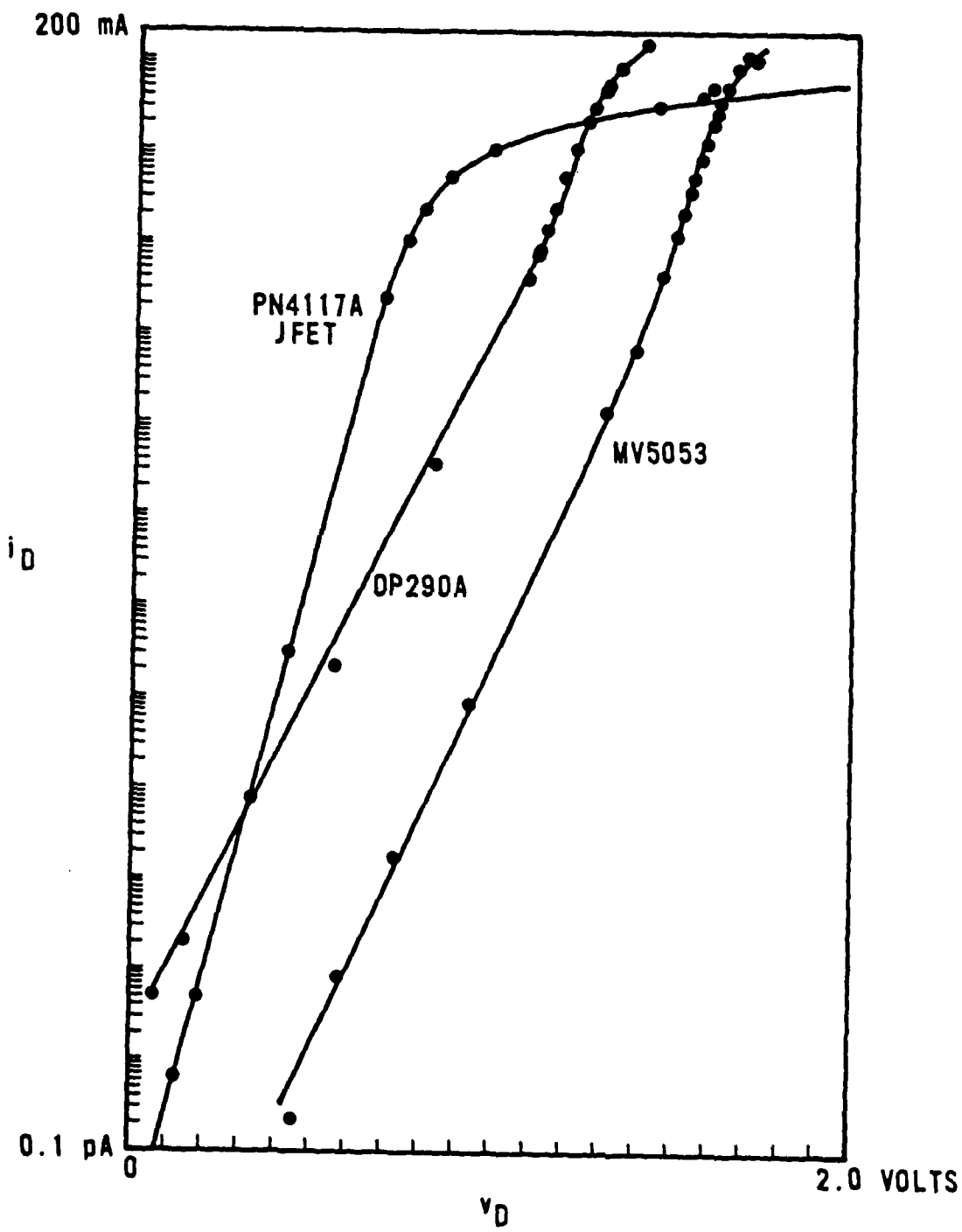
Fig.6-6

THYRISTORS

Semiconductor thyristors are PNPN structures that are useful for switching very large currents. The two most common types of thyristors are the silicon controlled rectifier (SCR) and the triac. The SCR can only conduct current in one direction. The triac can conduct current in either direction and is useful in AC applications. The SCR and triac each have two terminals that act like a normally-open switch, and a third terminal (which is called a "gate") that is used to turn the device on. The SCR is turned-on by injecting positive charge into the gate and out of the cathode terminal. The triac can be turned on by injecting either polarity of charge into the gate, however, triggering can be accomplished at lesser values of gate current when the gate and main terminal nr. 2 (MT2) has the same polarity.

When the SCR or triac is on, the voltage across the conducting diode is between about 0.7 and 2.0 volts, depending on the value of the current, with greater voltages for greater currents. Once the SCR or triac has been turned-on, it continues to conduct until the magnitude of the current in the device is decreased to less than the holding current, I_H. Values of I_H at a temperature of 25 celsius range between about 5 to 50 mA for devices that are rated to conduct 5 to 15 A_{rms}. Both the SCR and triac are readily available in models that can conduct rms currents on the order of 5 to 50 amperes, and are widely used as motor speed controllers and light dimmers.

There are several attractive features of thyristors for use in transient suppression circuits. Because the voltage across a conducting thyristor is

less than that for a zener or avalanche diode, or for a varistor, the
thyristor operates at a lower temperature for a given current. The thyristor
is also designed to routinely tolerate large surge currents (e.g. switching an
incandescent lamp on). A SCR or triac will conduct spontaneously (without
deliberately injecting charge into the gate) when the rate of change of
voltage across the device, dv/dt, is sufficiently great. Such conduction may
(and probably will) damage the SCR or triac, if the magnitudes of the anode
current and dI/dt exceed specified ratings for the condition of a marginal
gate trigger current. This bars many types of thyristors from most
applications in transient protection circuits.

The principal disadvantage to the use of SCRs and triacs as transient
suppressors is that during the turn-on period (typically about 1 or 2 μs) the
magnitude of dI/dt must be kept small. When the thyristor begins to conduct,
some regions begin to conduct before other regions. If dI/dt is too great,
the higher conductivity regions will develop into "hot-spots" with resulting
thermal failure of the device. SCRs have been developed that can tolerate
values of dI/dt of at least 200 A/μs. Asymetrical silicon controlled
rectifiers (ASCRs) are available that can tolerate values of dI/dt of at least
2 kA/μs. An ASCR needs to have a rectifier in series with the anode to prevent
conduction during reverse polarity.

One can decrease the turn-on time and increase the maximum tolerable
dI/dt by pulsing the gate of the SCR or triac with a very large gate current,
5 to 20 times I_{GT}. (I_{GT} is the minimum gate current that is required to
trigger the device when the main terminals are to carry the maximum steady-
state current.) Such gate triggering circuits are available in integrated

circuit form.

A possible disadvantage of SCRs and triacs is that once they are turned-on, they can be difficult to turn-off unless the current changes polarity regularly (e.g. 50 to 400 Hz AC power application) or naturally decays to zero (e.g. capacitor discharge appliction). In this respect they have the same problem as the power-follow of a spark gap that is in operating the arc mode. However, the holding current of a SCR or triac is usually an order of magnitude smaller than the extinguishing current for a spark gap.

The basic SCR transient protection circuit with a zener diode and resistor in series with the gate terminal, shown in Figs. 7-1 and 7-2, was patented by Gutzwiller (1965). De Souza (1967) added additional features, which will be discussed in Chapter 13 on protection of DC power supplies. Voorhoeve (1975) used a triac and bipolar zener diodes, as shown in Fig. 7-3, to reduce the number of components. These circuits have the disadvantage that as the thyristor conducts, the magnitude of the gate current decreases. This limited gate current increases the thyristor's vulnerability to damage from large values of dI/dt.

Chowdhuri (1974) suggested using a varistor in series with a SCR to turn off the SCR after the transient. This suggestion raises the question "Why not use a varistor alone?" The conducting SCR would have less than 2 volts across it and the varistor might have 300 volts across it (depending on the varistor model and value of the surge current). The varistor will absorb nearly all of the energy and essentially determine the clamping voltage. One concludes that there is little advantage in Chowdhuri's suggestion. If large clamping

BASIC SCR CROWBAR CIRCUIT
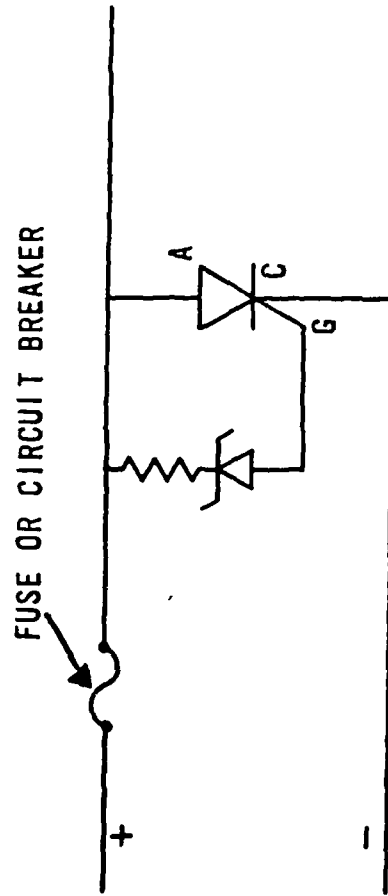
Fig.7-1

AC CROWBAR CIRCUIT (GUTZWILLER, 1965)
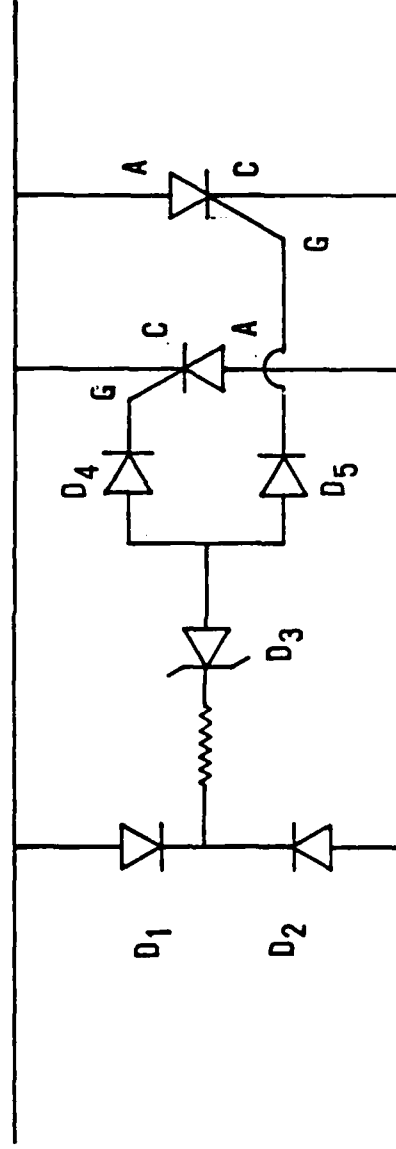
Fig.7-2

BASIC TRIAC CROWBAR CIRCUIT

FUSE OR CIRCUIT BREAKER

MT2

MT1

G
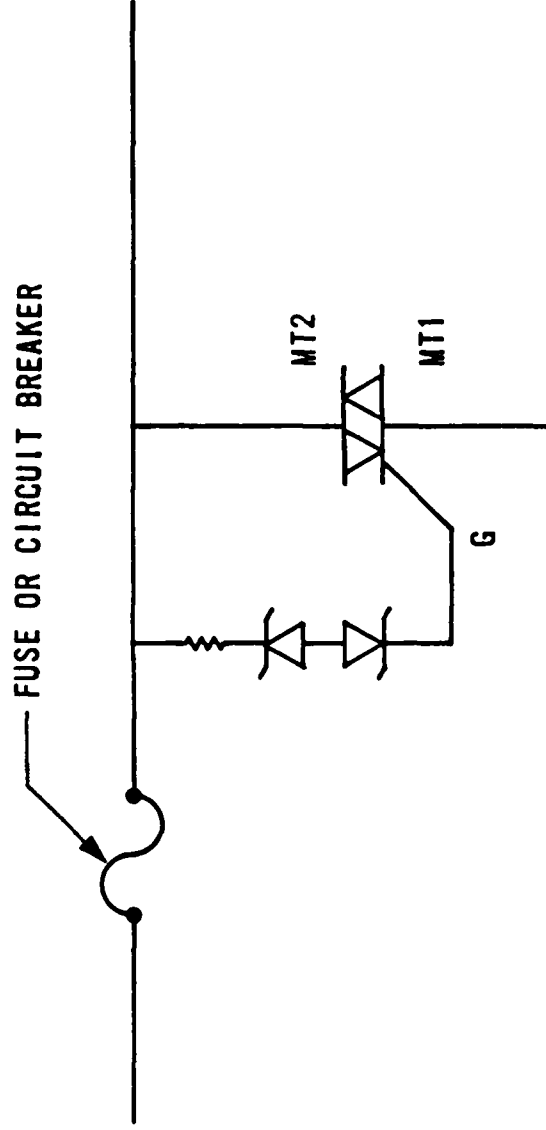
Fig.7-3

voltges are tolerable, a varistor alone can be used. Otherwise, a SCR circuit may be used, but a varistor in series with a SCR offers no advantage.

Also in the class of thyristors are many unusual devices such as (1) diacs, (2) silicon bidirectional switches, and (3) sidacs. Most of these devices are commonly used to supply gate current to a SCR or triac, and are known as "trigger devices." These devices tend to be non-conducting until a sufficiently large voltage of either polarity is placed across them, then they abruptly switch into a conducting state. The value of the voltage at which conduction (switching) takes place depends on the type of device and model number, but tends to be in the following range:

diacs	15 to 250 volts (30 volts common)
silicon bidirectional switch (SBS)	6 to 10 volts
sidac	95 to 250 volts

Manufacturers and part numbers for these unusual devices are listed below.

	diacs	SBS	sidac
General Electric	ST2	2N4991-4993	
Motorola	1N5758-5761	MBS4991-4992	MK1V series
Teccor	HT series		K1050E series

Integrated circuits are available to control a SCR for overvoltage protection. These integrated circuits provide a large gate current (more than

100 mA) and fast response (less than 1 µs) as well as offering a reduction in the number of components.

Motorola MC3423 and MC3424

Various integrated circuit protection devices exist that apparently incorporate an internal thyristor. This type of device was introduced in 1983 and is expected to become widely used for protection of telephone lines.

SGS-ATES	LS5120
RCA	TA9271 (developmental model)
Motorola	MPC2005
Thomson CSF	TPA100A-12, TPB100A-12, etc.

IMPEDANCES AND CURRENT LIMITERS (SERIES DEVICES)

INTRODUCTION

This chapter contains information on a variety of useful components in transient protection circuits. Some of these components (e.g. resistors, inductors, capacitors, and fuses) are familiar items to electronics engineers and technicians. We shall skip the familiar properties and concentrate on the specific details that are important in transient protection applications. Such details are often rather specialized knowledge. Other components (e.g. positive temperature coefficient resistors, bifilar chokes, lossy line) are specialized components that will be described more completely.

RESISTORS

Resistors are included in hybrid transient protection circuits, for example, between a spark gap and an avalanche diode. While this is not a precision application, it is critical that the resistor remain undamaged by exposure to transients. There are several ways that the resistor can fail to function as intended.

1. A spark discharge could form in the air surrounding the resistor (a phenomenon called "flash-over"). If the spark goes to ground, the electronic devices are not likely to be damaged. If the spark were to shunt the resistor, catastrophic damage to electronic devices downstream is possible.

2. Dielectric breakdown could occur inside the resistor package. The
resulting arc could shunt part or all of the resistive material, with a
resulting increase in current. The increased current may damage devices
that are downstream from the resistor.

3. Overheating produced by very large power dissipation could vaporize
the resistance material, and cause the resistor package to explode. If
the resistor were in series with the normal signal path, such an event
could cause the system to fail, just as if a wire were broken.

4. Overheating produced by very large transient power dissipation could
permanently damage the resistance material (e.g. cause a permanent change
in the resistance value of the order of 10% or more), without producing a
short or open circuit. It is likely that the resistor would continue to
be functional for transient protection applications after a few such
changes.

Consider a 22 ohm, 0.5 watt (steady-state) carbon composition resistor
that is connected between a 150 volt spark gap and a 15 volt avalanche diode.
If a 500 volt pulse were suddenly applied to the spark gap, the resistor would
have an instantaneous power dissipation of about 11 kW, before the spark gap
conducts. Most engineers who are unfamiliar with transient design would take
one look at the 11 kW power dissipation and conclude that the resistor would
be destroyed. If the spark gap were to remain non-conductive for 1 μs (a not
unrealistic assumption), about 11 mJ of energy would be deposited in the
resistor. The energy deposited in the resistor by the pulse is small, and it
is quite likely that the resistor would survive. Once the spark gap goes into

the arc mode (20 volts across the spark gap), the power dissipation in the resistor is only about 1.1 watts, a modest overload.

There have been several empirical studies of the ability of resistors to tolerate large pulse currents (Lennox, 1967; Domingos and Wunsch, 1975). The conclusion of this research is that carbon composition and wirewound resistors are able to survive large transient power dissipations. Metal-film and carbon-film resistors are not as suitable owing to concentration of current in a thin layer of material and owing to internal dielectric breakdown between adjacent turns of resistive material in a helix pattern.

A 100 Ω, 0.25 watt (type RN65) metal film resistor can survive a pulse with a 4 mJ energy content whereas a carbon composition resistor with the same ratings that is manufactured by Allen Bradley can survive 50 mJ (Lennox, 1967). Allen Bradley 1 watt carbon composition resistors between 51 Ω and 20 kΩ can survive pulses of at least 200 mJ.

In addition to the maximum power rating, which is familiar to all electronic engineers and technicians, resistors also have a maximum working voltage rating. The maximum working voltage rating is given as a steady-state value between the two terminals of the resistor for some common devices in the following list:

MIL style	DESCRIPTION	max working voltage
RC07	0.25 W carbon composition	250 volts
RC20	0.5 W carbon composition	350 volts
RC32	1.0 W carbon composition	500 volts
RC42	2.0 W carbon composition	750 volts

RW67V   6.5 W wirewound   400 volts

(e.g. Dale RS-5, IRC/TRW AS-5)

Operation of the resistor at greater voltages invites premature failure from internal dielectric breakdown. It is not clear how to apply this information to transient protection applications. Certainly, one can do a conservative design and specify resistors with steady-state voltage ratings equal to or greater than the worst-case transient situation.

The reader is cautioned that there is another maximum voltage given in resistor specifications, called "dielectric withstanding voltage". This value is the maximum voltage between either terminal and the exterior of the case of the resistor. The value of the dielectric withstanding voltage is important when the resistor is to be mounted next to a conducting metal chassis.

Real resistors have parasitic shunt capacitance and series inductance. Woody (1983) measured values between 1 pF and 2 pF for 17 different resistors of various types, with a typical value of 1.6 pF. At 10 MHz a capacitance of 1.6 pF has a reactance of about 10 kΩ, which is much larger than the resistance values that are commonly used in transient protection circuits. Thus, the parasitic capacitance of the resistors is not expected to be a problem in transient protection applications. Parasitic inductance is more significant in transient protection applications.

Woody (1983) measured the parasitic inductance of 15 different resistors. He found about 20 nH inductance in an Allen-Bradley 56 Ω, 0.25 W carbon composition resistor. Wirewound resistors had much larger parasitic

inductance: values ranged from about 0.3 μH to over 1 μH. Because transients are a high-frequency phenomenon, the inclusion of parasitic inductance in the real resistor will act to increase the magnitude of the impedance during transients. Since resistors are included in transient protection circuits to increase the impedance between two points (and thus limit the current), the parasitic inductance will usually be helpful in transient protection applications. For example, a parasitic inductance of 0.3 μH has a reactance of about 19 Ω at 10 MHz. If this inductance is in series with a 10 Ω resistance, the magnitiude of the impedance will be about 21 Ω, more than twice the resistance.

The idea of placing a resistor in series with a node to be protected is so simple that it is often ignored as a protective technique. A resistor alone is probably the least expensive transient protection circuit, although it has limited ability to protect vulnerable devices. Van Keuren (1975) showed that the minimum failure level for a 5510 integrated circuit line driver increased from 38 volts to more than 100 volts when a 270 Ω resistor was inserted in series with the input terminal. Burger (1974) briefly discussed protection of semiconductor junctions with a series resistor.

PTC RESISTORS

Nearly all non-linear devices that are used for transient protection of electronics have a smaller resistance for greater voltages across the device. Such devices are suitable for shunt elements in protection circuits. It would be attractive to also have a device whose resistance increased as the voltage across it increased. This kind of device would be suitable for series

insertion and would decrease the power dissipated in shunt protective elements that are located downstream from the non-linear series element. Unlike the shunt elements, the non-linear series element does not necessarily require a capability to carry a large current or absorb a large energy pulse. However, the series element must be able to tolerate large transient voltages across the device without breakdown or flashover.

There is a remarkable family of devices known as "positive temperature coefficient (PTC) resistors." The PTC resistors have an approximately constant resistance at temperatures below the switch temperature, T_S. At internal device temperatures above T_S, the resistance increases dramatically (e.g. $(1/R)(dR/dT)$ is 30 to 60 %/°C). The maximum resistance is typically a factor of 10^4 greater than the cold resistance (i.e. the resistance below T_S). Because of this dramatic change in resistance, PTC resistors are used as resettable fuses to protect motors or transformers from overheating or from damage by sustained excessive currents.

Most PTC resistors are fabricated from a bulk barium titanate semiconductor, $Ba\ Ti\ O_3$ (Andrich, 1966; Saburi and Wakino, 1963). Raychem Corp. fabrictes PTC resistors from conductive polymers. The polymer devices have carbon black dispersed in a plastic matrix. When the plastic reaches the switch temperature, it has a phase change, expands, and disrupts the conducting paths of carbon (Doljack, 1981).

The common values of T_S are between about 40 and 120 celsius, depending on the materials in the PTC resistor. Values of cold resistance range from about 22 Ω to 330 Ω for barium titanate devices and from about 0.01 Ω to 22 Ω

for conductive polymer devices.

The barium titanate PTC devices exhibit a negative temperature coefficient behavior when sufficiently large voltages are applied across the device. If such large voltages are maintained for a time on the order of one second or more, thermal runaway will occur with consequent degredation or destruction of the device. The designer of transient protection circuits should check prototypes of barium titanate PTC devices to be certain that after switching they maintain suitably large resistances at all voltage levels that could be encountered in the application. PTC devices that are rated for 120 V_{rms} (or more) maximum voltage appear to be suitable for insertion between a spark gap with a 90 to 150 V_{DC} firing voltage and a zener or avalanche diode with a smaller conduction voltage.

The conductive polymer PTC devices appear to be able to withstand very large voltages across the terminals without damage, and do not exhibit the negative temperature coefficient properties of barium titanate devices Moreover, the polymer PTC devices have a larger temperature coefficient than barium titanate devices, so the switching action of the polymer PTC devices is more abrupt.

PTC resistors are sold in the U.S.A. by the following companies:

Keystone Carbon Company, Thermistor Div.

Mepco/Electra (N.V. Philips components)

Midwest Components Inc.

Murata-Erie

Raychem Corp., Polyswitch Div.

Siemens

INDUCTORS

Because an inductor has a voltage drop that is proportional to the rate
of change of current in it, the inductor is an attractive device for transient
protection circuits. To be suitable for transient protection applications an
inductor should meet all of the following criteria.

1. Thick insulation on wire in the coil to prevent dielectric breakdown
between adjacent turns. Bell Laboratories (1975, p.108) recommends
magnet wire with a thick vinyl acetal resin varnish insulation, e.g.
Formvar, trademark of Belden Corportion. Regular hook-up wire with
polyvinylchloride or polyethylene insulation of at least 0.4 mm thickness
(and 1 kV rating) would probably be even more suitable, however it also
has a larger volume.

2. The high-permeability core in the inductor (e.g. iron or ferrite)
should not saturate at small values of transient current.

3. The inductor core should be insulated from the coil with TFE
(tetrafluoroethylene) tape [Teflon, trademark of du Pont] to prevent
both dielectric breakdown and removal of insulation from the wire by
mechanical stress (Bell, 1975, p.108). One might also consider polyvinyl
chloride insulated wire with a nylon jacket.

4. Multiple layers of wire on the coil should be avoided because this is inevitably associated with an appreciable capacitance between layers. This parasitic capacitance tends to short-circuit at least part of the inductor during transients.

5. Large diameter wire (e.g. at least 22 AWG) should be used for the coil winding due to mechanical stress expected during transients.

During surges with large values of current, the coil will encounter large mechanical stresses due to magnetic forces. If the coil is not properly designed, such forces can cause the wire to break. Miniature coils that are designed for use in radio-frequency circuits with steady-state currents on the order of 1 to 50 mA are not suitable for transient protection applications.

When inductors are inserted into transient protection circuits, there is a reduction in bandwidth due to the filter that is formed with other elements (e.g. parasitic capacitance of varistors, zener diodes, or avalanche diodes; deliberate inclusions of capacitors to form a filter). Digital and analog data communication lines often require a bandwidth of 10 kHz or more. Since lightning has appreciable energy in the frequency range between 100 Hz and 1 MHz, frequency discrimination is an inappropriate technique for protection against lightning when the signal bandwidth is in this region. Other transients, notably NEMP and ESD, have appreciable energy in the 1 MHz to 100 MHz region. Thus inductors are likely to be particularly beneficial when the signal bandwidth is restricted to either greater than about 500 MHz or less than about 20 Hz.

Another consideration is that for circuits that have a normal current on the order of a ten milliamperes or less, the IR drop across a series resistor of 10 to 50 ohms is not particularly objectionable. Most analog and digital signals are in this range of currents. Since resistors are less expensive than inductors, most circuit designers will specify a resistor rather than an inductor.

For these reasons, inductors have been more commonly used in transient protection circuits for power lines (DC to 400 Hz normal operation) than in analog or digital data lines. The large currents (e.g. 0.5 to 15 amperes) in power circuits make appreciable series resistance undesirable, so an inductor is the only acceptable series device. Furthermore, it is possible to design economical LC filters that offer considerable attentuation of transients, but have negligible effect on normal power transmission.

A "magnetic variable resistor", shown in Fig. 8-1, was invented by Person (1969) for use in a lightning surge arrester. Because the resistor is fabricated from a permeable material, it probably behaves more like an inductor than a resistor. Still, this is a clever device.

FERRITE

The inductance of a coil with a given shape and number of turns can be increased by placing a high-permeability core inside the coil or surrounding the conductor. For attenution of transients and noise, it is generally recommended that the core be a ferrite material. Ferrites are a ceramic

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS – 1963 – A

INSULATING SHEET

TERMINAL

CONDUCTING PERMEABLE STRIP

TERMINAL
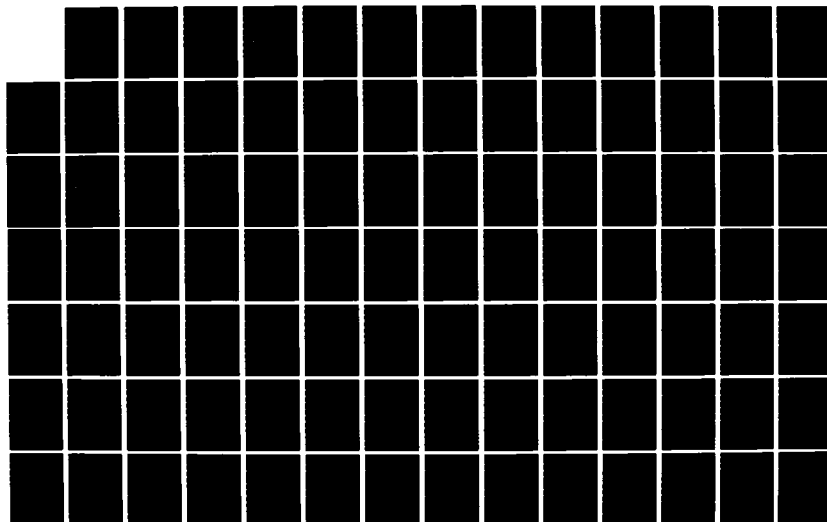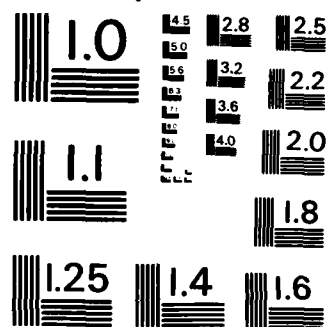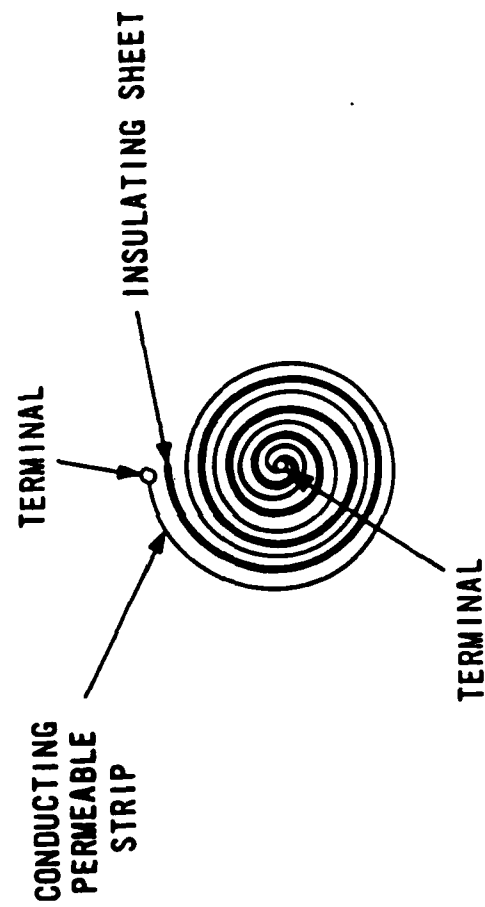
Fig. 8-1 MAGNETIC VARIABLE RESISTOR (H. R. PERSON)

material with a large permeability and small conductivity. At frequencies above about 1 MHz the effect of a ferrite core on a circuit can be modeled as a resistance in series with an inductance. The presence of the resistive (or lossy) part of the impedance will damp any oscillations caused by the interaction of the inductance with capacitance elsewhere in the circuit. In addition, the resistive part of the impedance will act to dissipate transients rather than just reflect them (Bell, 1975, p.106-7).

Small cylinders of ferrite material with an axial hole for conductor(s) are called "ferrite beads" and are widely used to cure electromagnetic interference problems. As an indication of the effect of ferrite beads, a representative specimen was slipped on a piece of 20 gauge copper wire that was fashioned into a loop with a diameter of about 9 cm. The impedance of the loop was measured on a Hewlett-Packard model 4192 LCR meter before and after the bead was inserted. At 100 kHz, a single Indiana General (Permag) F1832-1-06 bead added about 7 µH to the loop.

BIFILAR CHOKE

A bifilar choke is a four terminal component that is useful for inserting inductance in series with common-mode sources, but has negligible inductance in series with differential-mode sources. In this way, the use of a bifilar choke on a balanced line has a minimal effect on the bandwidth for normal signals, which are differential-mode, but offers substantial attenuation of common-mode transients and noise. All bifilar chokes have two independent coils of the same size and number of turns. These coils are commonly wound on a toroidal core of ferrite, as shown in Fig. 8-2. The use of a ferrite core

BIFILAR CHOKE
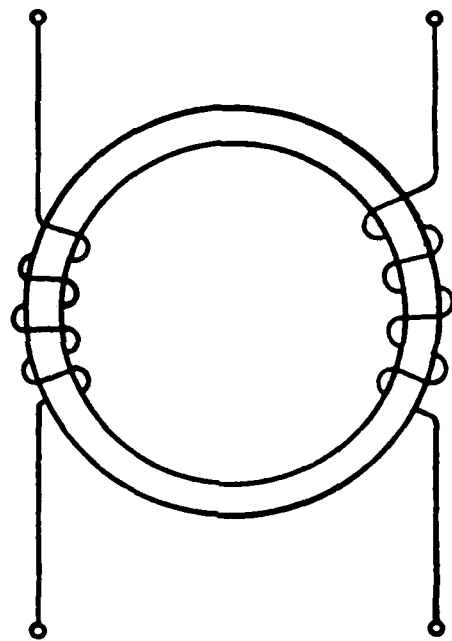
Fig.8-2

introduces an appreciable resistive (lossy) component to the reactance, which is a desirable feature for transient and noise suppression. An additional feature of bifilar chokes is that there is no core saturation from large differential-mode currents (Sandia, 1972, p.348).

Measurements were made of the common-mode and differential-mode inductance presented by two commercial bifilar chokes on a Hewlett-Packard model 4192 LCR meter. These chokes are wound on a toroidal ferrite core.

Bifilar Choke	frequency	common-mode inductance	diff-mode inductance
Siemens B82722-G2-A10	0.1 MHz	9.7 mH	29 µH
	1.0 MHz	16.6 mH	31 µH
Siemens B82724-G2-A14	0.1 MHz	8.4 mH	25 µH

Notice that the differential-mode inductance is about 0.3% of the common-mode inductance at 100 kHz. Therefore, these bifilar chokes will present a larger series impedance to common-mode signals than to differential-mode signals.

Bifilar chokes are one of the oldest transient protection techniques. A relatively ancient reference to bifilar chokes is in a patent by Rovere, Estes, and Milnor (1934). These engineers were interested in protecting carrier telegraph and telephone circuits from lightning.

CAPACITORS

Capacitors are commonly used as shunt elements in low-pass filters and as bypass applications on DC supply lines. Capacitors may also be used as series

elements when the signal frequency is greater than about 10 MHz (e.g. radio frequency applications). The important properties of capacitors in such applications are capacitance (of course), DC voltage rating, the parasitic inductance of the real capacitor, and the details of what happens when the voltage rating of the capacitor is exceeded.

An ideal capacitor has an impedance that is proportional to $1/f$, where f is frequency. However, for all lumped element capacitors (the kind used in circuits below a few hundred megahertz), there is a resonance frequency f_0 due to internal inductance of the capacitor. At frequencies greater than f_0, the impedance of the capacitor is dominated by the inductance term. The impedance is then proportional to f, instead of the $1/f$ proportionality that one would expect of an ideal capacitor. The value of f_0 is usually between 1 MHz and 100 MHz for non-electrolytic capacitors that are commonly used in electronic circuits. This implies that a capacitor in a circuit may behave as an inductor when transients with appreciable high-frequency content are present. Techniques for reducing the effects of inductance in capacitors on circuit performance is discussed in the Chapter 9 on parasitic inductance.

Tasca (1981) described tests of various capacitors when charged with a pulse of constant current. Breakdown voltages of glass, mica, ceramic, plastic, and paper dielectrics were independent of the rise time of voltage over the interval between 0.01 µs and 20 µs. For capacitors with a DC breakdown voltage rating of 100 to 600 volts DC, the actual breakdown voltage was usually about a factor of 10 to 15 greater than the rated voltage. The ratio of actual breakdown voltage to the rated voltage gives a safety factor. About 0.3% of the capacitors had a safety factor as small as 2. This provides

a comfortable margin of safety.

Tasca (1981) found that tantalum electrolytic capacitors when pulsed with a polarity that was the same as the normal working polarity exhibited a breakdown behavior similar to an avalanche or zener diode. The voltage across the tantalum capacitor was approximately constant at a value between 2 and 10 times the maximum rated working voltage.

When capacitors are used in applications where the voltage rating may be exceeded, it is important to specify "self-healing" capacitors. Such capacitors are commonly made with metal electrodes that are deposited in vacuuo on a plastic dielectric sheet. (If the dielectric, for example, is polyester, the capacitors produced by this method are known as "metalized polyester" types.) The thickness of the electrode is usually less than a few micrometres. If the dielectric should break down, the resulting arc evaporates the thin electrode in the region of the arc. The arc then extinguishes automatically when the radius between the dielectric puncture site and the solid metal electrode becomes too great to maintain the arc. The capacitor is not damaged by this process, which is known as "self-healing".

Metalized plastic dielectric capacitors are not the only self-healing capacitors. One can also make self-healing capacitors with a paper dielectric. Paper dielectric capacitors are obsolete in essentially all applications except AC power line bypassing, an application that is relevant to transient protection.

It may seem plausible to use a shunt capacitor to store the charge that

is transferred during electrical overstress, thus reducing the voltage. This technique can be quite effective for electrostatic discharge (ESD), where the total charge transfer is usually less than one microcoulomb. We get this value from considering the 150 pF capacitance of the human body at 6.7 kV; 6.7 kV is a larger voltage than usually encountered during ESD. The addition of a 0.01 μF shunt capacitor will reduce the 6.7 kV potential difference to slightly less than 100 volts.

However, when large transients from lightning or NEMP are encountered, it is not economical to divert the transient with a shunt capacitor. For example, the standard 8×20 μs test waveform with a peak current of 10 kA contains 0.328 coulombs of charge. A capacitance of 3280 μF would be required to have 100 volts across the line. This large value of capacitance is economically obtainable only when aluminum electrolytic capacitors are used. Such capacitors have relatively large internal series inductance and are not effective for transient suppression or pulse current applications. Even if one could obtain a suitable capacitor, the drastic reduction in system bandwidth and increase of output current drive requirements that results from insertion of such a large capacitance value would probably be unacceptable except on DC power supply lines. If the output impedance can be kept to 1 Ω, the system could still have a 48 Hz bandwidth, which is adequate for many slowly-varying signals. However, the amplifier whose output is connected to this 3280 μF capacitor must supply a current with a peak value of 10 amperes, in order to obtain a 10 volt sinusoidal signal at 48 Hz. Typical output currents of integrated circuit amplifiers are of the order of 10 mA, so one would have to sacrifice either the amplitude of the output voltage or bandwidth in order to use these integrated circuit amplifiers with such a

large capacitance.

FILTERS

Prefabricated filters are one of the principal components available to engineers to prevent systems from dumping high frequency ($v > 0.5$ MHz) noise into power lines and into communication cables (or to prevent such noise from entering a system on the same conductors). For optimal benefit, these filters must be used in conjunction with an effective shielding technique.

In order to save space and maintain the integrity of shield topology, filters are being included in pins inside connectors. The mating connector is attached to a conducting plane, commonly a bulkhead or chassis. The noise removal must be accomplished in a feed-through arrangement as shown in Fig. 8-3. Fig. 8-4 illustrates the two other ways to connect a filter, both of which are undesirable. There are no other good choices given a single layer shield.

Before one attempts to use a filter module for attenuation of electrical overstresses, one must be certain that the spectrum of the possible overstresses does not overlap the bandwidth of the signal. A condensed review of the nature of overstresses was given in Chapter 1. We saw that overstresses can have nanosecond rates of rise (e.g., electrostatic discharge and EMP from nuclear weapons) and can continue for tens of milliseconds (e.g., continuing current in lightning), the possible spectrum for overstresses extends from about 100 Hz to 1 GHz. However, not every possible source of overstress will be important in a particular application, so that, for a particular application, the threat spectrum might be considerably smaller than

Fig.8-3

"CLEAN" ENVIRONMENT

NOISY ENVIRONMENT

FEED THROUGH

METAL SHIELD
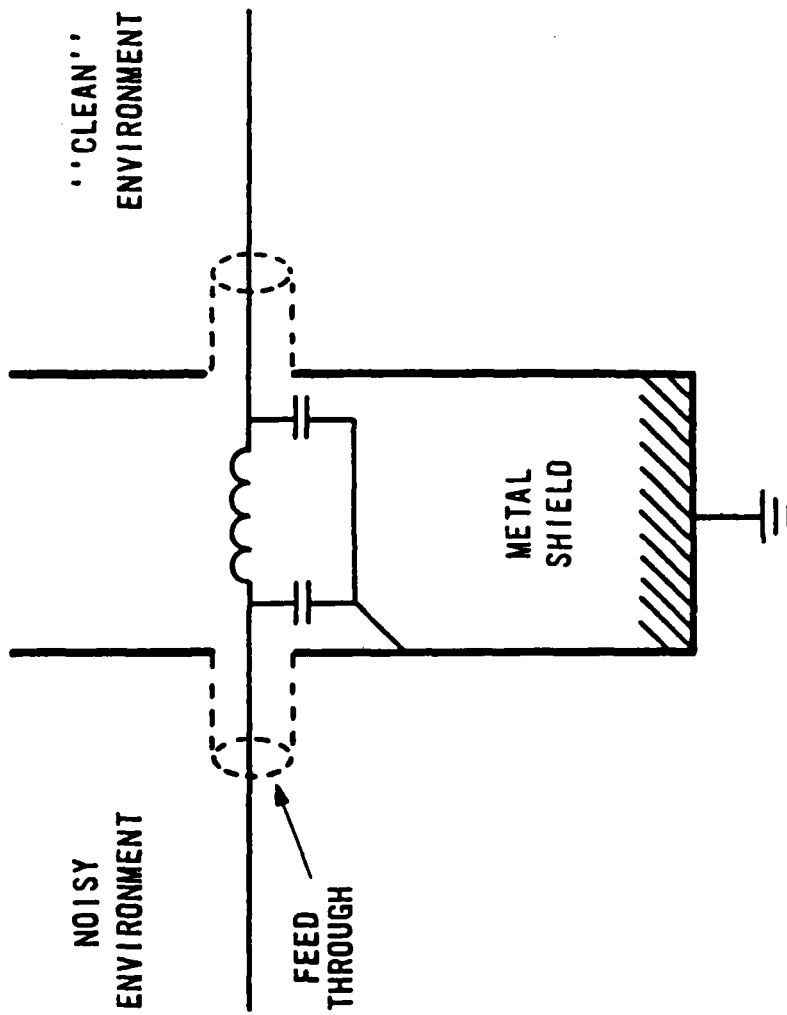
Fig.8-4 TWO BAD WAYS TO USE A FILTER TO REMOVE NOISE FROM A LINE

NOISY
ENVIRONMENT

CLEAN
ENVIRONMENT

NOISE

NOISE

NOISE

SHIELD
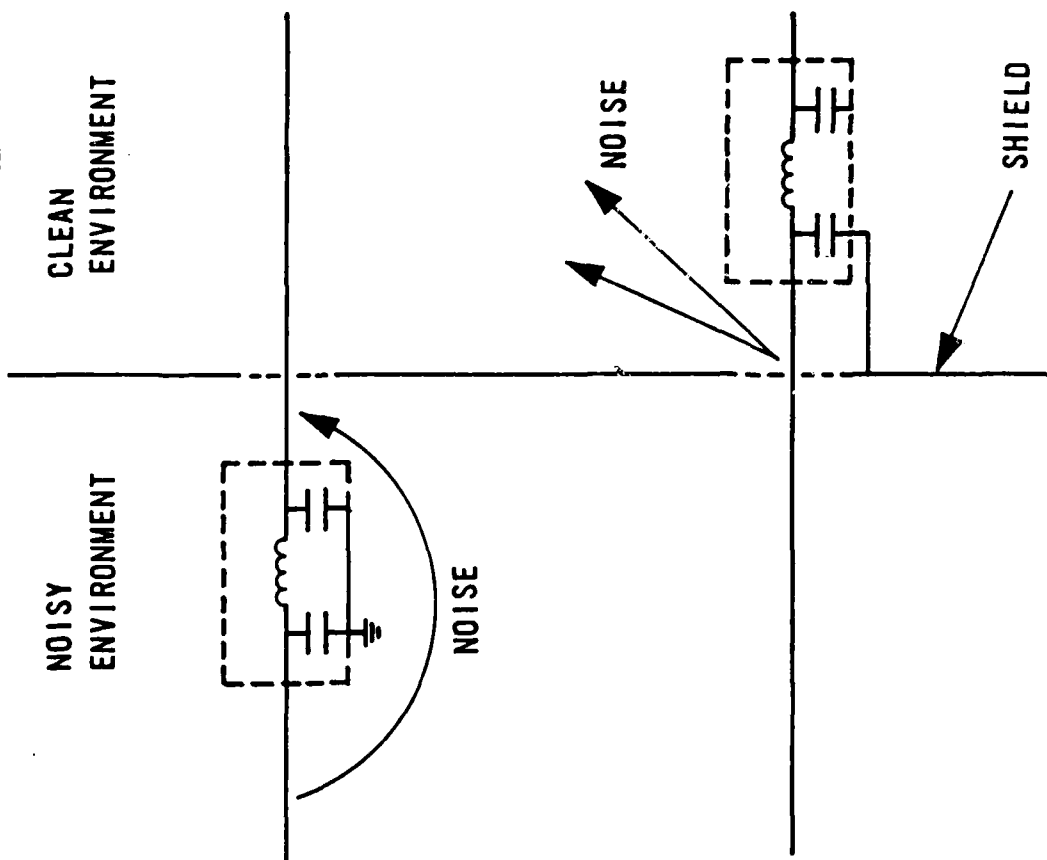
this estimate. If the overstress spectrum includes the bandwidth of the signal, then filters are not a reasonable protection technique.

In some situations conductors can carry large amplitude transients (currents greater than 100 A, voltages greater than 2 kV). It is not economical to design filters for such situations when non-linear shunt elements (e.g. spark gaps and varistors) are available.

The small shunt impedance of a spark gap, varistor, diode, or other non-linear device during a transient can cause reflections. In this view, the use of a non-linear device did not "solve" the problem of transients, it merely dumped the transient back on the line. This is frequently mentioned as a reason not to use non-linear devices. However, low-pass LC filters (without lossy elements) also reflect transients. While we can hardly advocate casually introducing reflections, one should recognize that reflecting the transient away from a vulnerable load does protect the load.

The application (and particularly mis-application) of non-linear devices is not without hazards and disadvantages. However, reliance on filters (and shielding) alone can also be unsatisfactory. In particular, the very large voltages and currents in some worst-case transients overwhelm filters. Laboratory tests have shown that some prefabricated filter modules suffer arcing between an input terminal and ground at voltages between 5 kV and 10 kV. Failure of insulation can also occur inside the module, where it may cause degredation of capacitors or inductors.

A spark in air at the external terminals of a filter module has

approximately the same effect on the circuit as a discharge inside a spark
gap. Actually the spark gap is better: its properties are more predictable,
and the spark gap may respond more quickly owing to radioactive preionization
and other features. Aboard a military aircraft, for example, the air
pressure, and hence the breakdown voltage, changes dramatically with altitude.
The spark gap, which was discussed in Chapter 3, is a sealed system that is
free of effects due to dust, dirt, and ambient air pressure.

Breakdown of insulation inside the filter module is generally
undesirable. If it can be confined to "self-healing" capacitors, the
insulation breakdown may have no serious consequences for the reliability of
the filter module. However, it is not simple to predict the output of the
filter module during internal breakdown.

Given that filters fail by either internal or external arcing, one may as
well combine spark gaps and/or varistors with filters that may be exposed to
high voltages. In that way the curent is carried in a predetermined path and
does not degrade the filter's future performance. This recommendation has
also been made in DoD DCPA TR-61A (1977). Bell Laboratories (1975, p.107-8)
recommends that filters have an inductive input and be preceded by a spark
gap, as shown in Fig. 8-5. The input inductor of the T filter shown in Fig.
8-5 is said to be less vulnerable to damage than the input capacitor of a π
filter. Moreover, the input inductor produces a large voltage across the
spark gap during a transient that has a large value of dI/dt. This large
voltage across the spark gap causes the spark gap to conduct more rapidly.

The internal breakdown problem can be eliminated by installing

PROTECTED
PORT
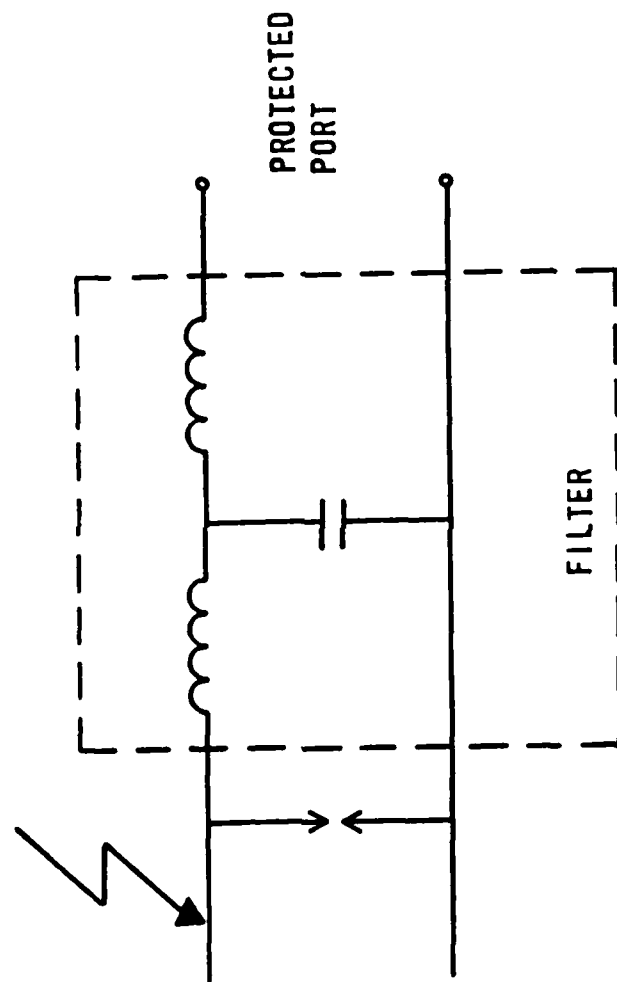
FILTER

Fig.8-5

appropriate non-linear devices in parallel with shunt capacitors in the filter. Metal oxide varistors would be particularly desirable in this application since they can conduct large surge currents, have no discontinuity in their (V,I) curve, and respond rapidly to overvoltages. In addition, one can easily obtain varistors that have negligible conduction (less than 1 mA DC) at voltages below 1 kV. The parasitic capacitance of varistors is large enough (100 pF to 10 nF) to act as an additional shunt capacitance, which might act to reduce the effects of parasitic inductance in the filter capacitor, or even replace the filter capacitor.

The concept of combining varistors and ferrite core inductors to form a π filter, as shown in Fig. 8-6, has been investigated and endorsed for filter modules for signal lines (Campi, 1977). Campi's filter was to be a pin in a multi-pin electrical connector. In this way, the transient protection was conveniently included in the system along with proper shielding.

Prefabricated filter modules are characterized in a test fixture with purely resistive 50 Ω source and load impedances. However, in actual use the source or load impedance might be reactive. In such a case, a resonant circuit might be formed, which might amplify, rather than attenuate, the transient (Bridges, et al., 1976, p.6-5). As in many other situations, design of equipment requires confirmation in the laboratory before the equipment should be put into production.

LOSSY LINE

Most transient protective devices operate by shunting the transient to

CONNECTOR

FERRITE
BEAD

MOV
SLEEVE

INDUCTOR

METAL HOUSING

FERRITE
BEAD

CONNECTOR
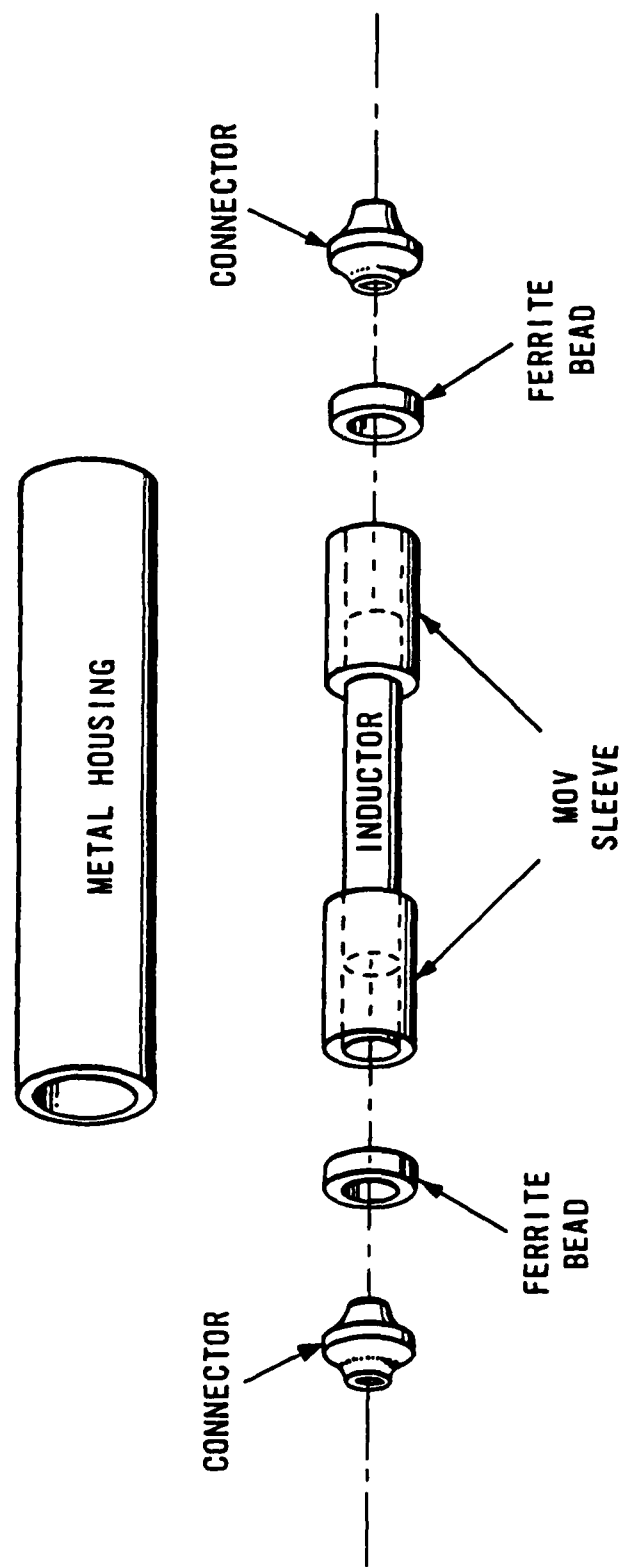
Fig.8-6 CAMPI'S FILTER PIN PROTECTOR

ground through an element with a very small impedance (e.g. spark gap, varistor, diode, or capacitor in a π filter). When the interaction of the line and the transient protection circuit is viewed from transmission line theory, one realizes that an appreciable part of a rapidly rising transient will be reflected by the TPD which approximates a short-circuit termination. Other types of transient protection circuits (e.g. the inductor input of a T filter) may approximate an open-circuit termination, again producing a reflection of an appreciable part of a rapidly rising transient.

It would be desirable to attenuate these reflections. Otherwise, the reflections may radiate energy into other systems and cause problems there. Also the reflections may produce overstress on the cable and cause insulation to fail. Ultimately, the transient energy must be converted to heat. We want to do this in a controlled way so that no damage occurs anywhere. One could use ferrite core inductors or resistors to accomplish this.

There are several ways in which the cable itself can be converted into a filter to exponentially attenuate transients that propagate along the cable. The simplest way is to use wire with a large resistance per unit length. This is the solution adopted by manufacturers of oscilloscope probes for use with a (1 MΩ || 20 pF) input impedance. The coaxial cable in these probes has a center conductor of nickel-chromium alloy and a resistance per unit length between about 130 Ω/m and 350 Ω/m. Unfortunately for transient protection applications, this resistance wire only has a diameter of the order of 75 to 100 μm. During a large current pulse this thin resistance wire might melt.

A more robust coaxial cable with a center conductor of resistance wire is available as military type RG-222 (formerly RG-21A). The center conductor is 1.4 mm diameter Nichrome wire, the outer conductor is a double shield of silver plated copper braid.

$$C/\ell = 95 \text{ pF/m}$$

$$L/\ell = 0.24 \text{ }\mu\text{H/m}$$

$$R/\ell = 0.65 \text{ }\Omega/\text{m}$$

$$Z_0 = 50 \text{ }\Omega$$

-0.14 dB/m at 10 MHz

-0.31 dB/m at 50 MHz

-0.42 dB/m at 100 MHz

-0.60 dB/m at 200 MHz

-0.87 dB/m at 400 MHz

Another approach is to use a magnetic lossy medium between the center and outer conductors of coaxial line. This approach has been used by Capcon, Inc. in their proprietary "LossyLine" cable. They report attenuations of 65 dB/m at 100 MHz with this cable. Another proprietary Capcon, Inc. cable uses a helical center conductor with lossy material both inside the helix and between the helix and the outer shield. This helixed cable has a reported attenuation of about 425 dB/m at 100 MHz.

All cable will become lossy at very high frequencies owing to skin effects. At high frequencies the current is concentrated in a very thin layer, called "skin", on the exterior of conductors. The thickness of the

skin is inversely proportional to the square root of frequency, so as the frequency increases the skin depth decreases. As the skin thickness decreases, so does the cross sectional area for current flow. Since the conductivity of the metal is independent of frequency, the resistance per unit length increases with frequency. As a consequence, cables become more lossy at high frequencies.

FUSES AND CIRCUIT BREAKERS

The conventional wisdom is that fuses and circuit breakers are too slow to be effective in transient protection circuits. However, there are several reasons why we ought to discuss them anyway. First, generalizations that are usually valid have a way of inhibiting progress by restricting designers from breaking the rule in exceptional cases. Second, fuses continually appear in transient protection circuits, often in ways that may do more harm than good. Third, there are valid reasons to include fuses in the output of DC power supplies and to isolate defective loads from an AC power line. Thus fuses might be included in a transient protective circuit, with resulting increase in reliability of the system.

There are many different types of fuses and circuit breakers, each of which is available in a variety of different current ratings. A fuse is an expendable component that is designed to become a permanent open-circuit when exposed to excessive current. A circuit breaker is a device that opens a circuit when excessive current is detected, but can be reset manually (or automatically on some models) to restore the circuit.

The common model of "fast" fuse used in electronic equipement is a glass cylinder 1.25 inches in length, 0.25 inch diameter (known as type 3AG or AGC). The 1 A rated model typically takes 1 second to open at a 2 A current (maximum of 5 seconds at 2 A). At very large overloads the fuse is much quicker: at 10 A the 1 A fuse typically takes between 7 ms and 20 ms to open (Buss catalogue SFB, c.1978; Littelfuse Catalogue 19, 1977; Gould Shawmut Bulletin GF, 1980).

Circuit breakers tend to be slower than fuses, although they are satisfactory to prevent fires from excessive current in wires.

Martzloff (1983) found that the value of the action integral, $I^2 \Delta t$, required to open a fuse was essentially constant from 0.1 seconds to about 10 µs. A "fast" 10 A fuse had $I^2 \Delta t = 200$ A^{2}s. To open this fuse during a 8×20 µs waveform, a peak current of about 10 kA was necessary. This is near the maximum rated interrupting capability of this fuse. At larger peak currents the arc inside the fuse that is opening may not extinguish.

There are two common ways to use a fuse as shown in Fig. 8-7. Before we discuss the circuits shown in Fig. 8-7, it is worthwhile to explictly mention that the fuse shown in Fig. 8-7 is symbolic of any device that normally has a low impedance, but changes to a very large impedance when excessive current passes through it. In particular, the fuse could be replaced by a PTC resistor or a circuit breaker. Likewise, the varistor shown in Fig. 8-7 is symbolic of any device that normally has a large impedance, but changes to a very small impedance during electrical overstress. In particular, the varistor could be replaced by a spark gap, avalanche diode, or triac.

Fig.8-7

POSSIBLE OTHER SYSTEMS

RED INDICATOR LAMP

SOURCE

FUSE

TPD

(a)
LOAD TO BE
PROTECTED

(a)

SOURCE

FUSE

TPD

(b)
LOAD(s) TO BE
PROTECTED

GREEN
INDICATOR
LAMP

(b)

SOURCE

FUSE

TPD

RED
INDICATOR
LAMP

(c)
LOAD(s)
TO BE
PROTECTED

(c)
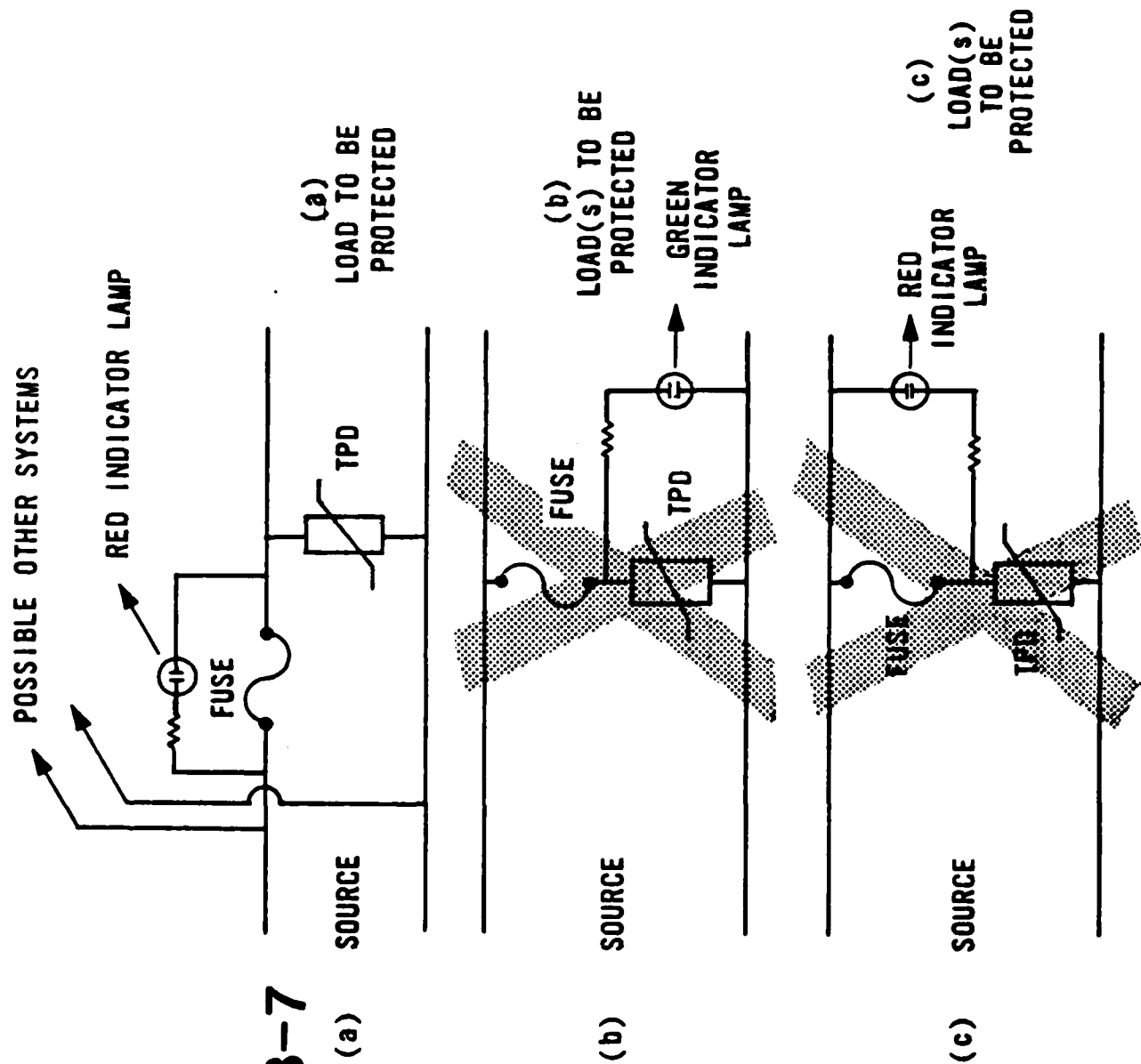
There are several possible hazards that could result in damage to the source in Fig. 8-7 : (1) the load may fail and draw excessive current, (2) a crewman or repairman may make a mistake while connecting cables and inadvertently short-circuit the source, (3) the insulation on the wires may be damaged and short-circuit the source, or (4) the varistor might fail as a short-circuit. The circuit in Fig. 8-7a protects against all of these hazards to the source, while the circuits in Fig. 8-7b and 7c only protect the source from a shorted varistor. The circuits in Fig. 8-7b and 7c do not provide comprehensive protection for the source. Therefore, one must use the circuit of Fig. 8-7a when two or more critical and independent systems are to be operated from the same source.

In some critical systems it may be intolerable to risk failure of the system due to interruption of power by a blown fuse. In this situation, the circuit of Fig. 8-7a can not be used. The circuit of Fig. 8-7b is often recommended because it can not interrupt power to the critical load. The circuit of Fig. 8-7b also has the advantage that it will jettison the varistor if the varistor draws excessive current and threatens the power supply or load. While many engineers have endorsed the circuit shown in Fig. 8-7b, it has some major shortcommings, many of which are not obvious.

First, the resistance of the fuse will increase the transient voltage across the power line, and thus acts to partly defeat the purpose of the varistor. This is not a trivial issue: a 1 A fuse has a resistance between about 0.1 Ω and 0.3 Ω (depending on ambient temperature and current). At a transient current of 1 kA, this resistance of the fuse contributes an extra 100 to 300 volts of overstress across the load.

Second, the parasitic inductance of the fuse can increase the transient voltage across the power line. The standard glass fuse and a minimal length of connecting wire has a length of about 5 cm. This gives an inductance of the order of 50 nH. At a value of dI/dt of 2×10^9 A/s, the extra overstress across the load due to the inductance of the fuse is about 100 volts.

Third, if the fuse blows during a transient when the current is rapidly changing, the sudden interruption of the transient current can create a very large voltage spike owing to the inductance in the wiring. This effect has been observed by M. N. Smith (1973, pages 5,9). Smith commented that the transient that was produced by the opening of the fuse was more severe than the transient that was applied to test the protection circuit.

Fourth, once the fuse opens, the system is without transient protection. During a thunderstorm or nuclear war the system might survive for only a few tens of seconds before the next transient destroyed the unprotected system.

If reliability of the system is indeed paramount, as was asserted during the rejection of Fig. 8-7a, one should ask "Do transients really threaten the system?" If transients are a threat, then provide a varistor that is more than adequate to withstand the worse-case transient overstress and omit the fuse. If transients are not a threat, then the varistor and the fuse are both superfluous (but the varistor alone might be included "just in case").

If the circuit of Fig. 8-7b is to be used in spite of its shortcommings, it would be prudent to include an lamp as shown in Fig. 8-7b to indicate that

the varistor is still connected to the circuit and providing protection. Because the lamp is on during a "good" condition and off during a "warning" or "failure" condition, the lamp should have a green color, not a red or orange color. Miniature fluorescent glow lamps are available in green; long-life tungsten lamps (with a green lens) or green LEDs are also possibilities. A disadvantage of this scheme is that a lamp that signals failure by turning off does not naturally attract attention, and is therefore poor human factors engineering. Further, in applications where the varistor is required to be installed in a remote location, "extra" cabling will need to be provided to route current to the lamp, since the lamp must be located in an area where it is readily observed. Replacement of the fuse in such a situation might be a major maintenance task. (If the fuse were to be located in a convenient place, far from the varistor, the inductance in the cable to the fuse would vitiate the transient protection offered by the varistor.)

The circuit of Fig. 8-7a, however, can use a neon lamp and a series resistor (of the order of 47 kΩ, 1 W) to display a bright red light to signal failure of the fuse. This lamp may be mounted far from the fuse without compromising the operation of the circuit.

The circuit of Fig. 8-7c is a particularly bad one. This circuit has the same undesirable features as Fig. 8-7b, since the fuse and TPD are connected in the same way. The only difference between Figs. 8-7b and 8-7c is in the way the indicator lamp is connected. In the circuit of Fig. 8-7c the indicator lamp is lit when the fuse has opened <u>and</u> the TPD has shorted. This would seem to be a good feature, since the lamp calls attention to the defective protection circuit. However, suppose the transient overstress blows

the fuse but does not short the TPD. Then the TPD will not be connected across the line (the resistance and inductance of the tungsten lamp vitiates the protection), but the lamp will not indicate this failure state. Alternately, suppose the circuit operates "properly." A large overstress shorts the TPD and blows the fuse. The light then comes on. However, the next large transient will cause the lamp to become an open-circuit, since there is no longer any transient protection in the circuit. If the crewman did not notice (and remember) that the lamp was on briefly, the system would remain without transient protection. This circuit has been developed for use by the military (Reynolds, 1972). It would be prudent to avoid this circuit.

ISOLATION DEVICES

INTRODUCTION

An isolation device is placed in series between the source and load, as shown in Fig. 9-1. An isolation device has no conductive path between input and output ports, this defines "isolation." It may seem queer to discuss an electrical component that is used in a series circuit, yet has no conductive path. Yet such devices can be quite useful to block common-mode voltages, shown as v_C in Fig. 9-1, from appearing across a load.

There are two common ways that the signal can be coupled from the input to output port:

       1. *magnetic field* (device is called an "isolation transformer")

       2. *light beam*   (device is called an "optical isolator").

One can also use an acoustic transformer: the electrical input signal is converted to sound waves with a piezoelectric transducer, the sound waves are converted back to electrical signals with a microphone. This clever scheme may be sensitive to ambient acoustical noise and will not be discussed further. We shall discuss isolation transformers first, then optical isolators.

ISOLATION TRANSFORMERS

In elementary theory, a transformer is a differential device. That is, the output (secondary) voltage is only a function of the voltage difference across the input (primary) terminals. In reality there is some capacitance between the primary and secondary coils, so that the common-mode voltage at
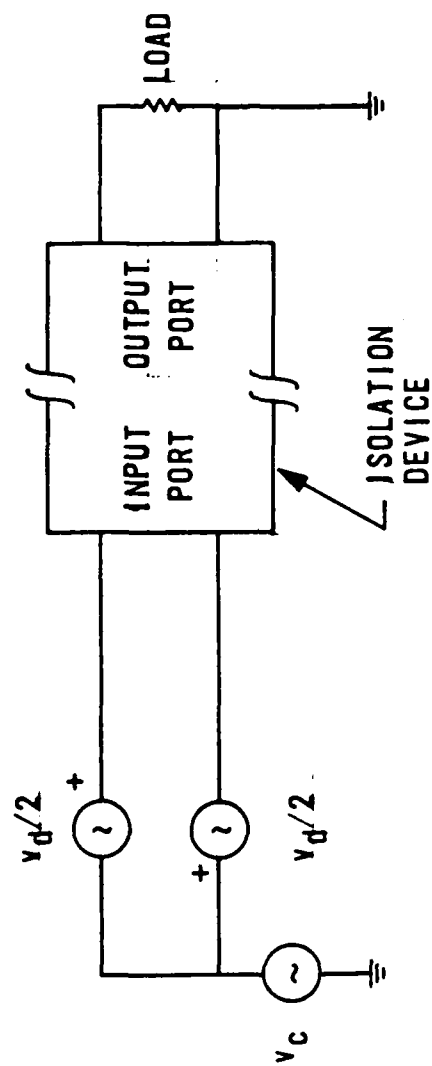
Fig.9-1

the input is coupled to the output. By interposing one or more electrostatic
shields between the primary and secondary coils, the effective capacitance
between the input and output ports can be reduced to negligible values. When
we do this, we obtain an isolation transformer, as shown in Fig. 9-2. A good
isolation transformer will have less than 0.005 pF capacitance between the
input and output terminals. If the load, which is connected to the output
terminals, has a 0.01 µF shunt capacitance, we have a common-mode attenuation
of 126 dB by simple voltage division.

Isolation transformers are most commonly used to block common-mode
voltages on the mains, an application that will be discussed later in Chapter
12 on AC power applications. Note that the ideal isolation transformer
provides NO differential-mode attenuation. Therefore, a differential-mode
voltage transient that appears at the input will be transmitted to the output
side. Also note that an ideal isolation transformer provides no voltage
regulation. If the input voltage drops to 95 volts rms, so will the output
voltage. There are techniques to add differential-mode transient attenuation
and voltage regulation to an isolation transformer. When we do this, the
resulting product is no longer just an isolation transformer, but is now
called a "line conditioner." We shall describe line conditioners in the
chapter on AC power applications.

OPTOISOLATORS

An optoisolator is an electronic component that contains a light source
and a photodetector, with no electrical connection between the two. A light
beam transfers information from the input to the output. Electrical

ISOLATION TRANSFORMER CONSTRUCTION

BOX SHIELD
COMPLETELY
ENCLOSES
SECONDARY
COIL

SHIELD IN
IRON CORE
BETWEEN
TWO BOX
SHIELDS

BOX SHIELD
COMPLETELY
ENCLOSES
PRIMARY
COIL
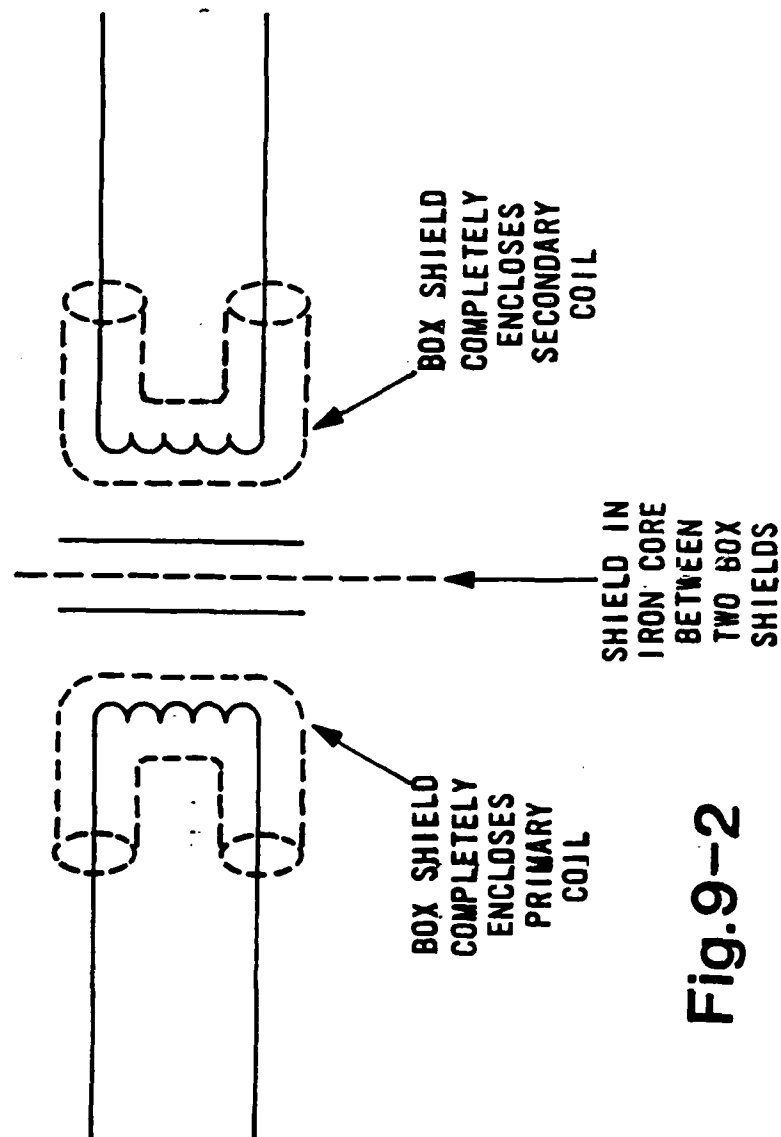
Fig.9-2

insulation between the light source and detector is provided by a piece of transparent glass or plastic. This insulation can typically withstand several thousand volts continuously.

The light source in nearly all modern optoisolators is an infared light emitting diode (LED). The photodetector in modern optoisolators is usually a silicon phototransistor, but silicon photodiodes and light-activated SCRs or triacs are also available. One particular photodetector, the photo-Darlingon deserves special mention in connection with transient protection and upset prevention circuits. The photo-Darlington is quite slow: response times of 0.1 ms to 1 ms are common. This inherent slowness provides immunity to transients that have too brief a duration to affect the output state.

The capacitance between input and output ports of an optoisolator in a six pin dual-inline-package (DIP) is typically between 0.3 pF and 2.5 pF. This capacitance and the shunt capacitance across the output port determines the common-mode rejection of the optoisolator. While a input-output capacitance on the order of 1 pF may seem small, at a 1 kV/μs rate of change of common-mode voltage, a current of 1 mA will be transfered from the input to output due to this parasitic capacitance. Any attempt to intercept this current with a guard ring will inevitably decrease the spacing between the input and output circuit connections and thus reduce the isolation voltage. When large rates of change of common-mode input voltage are anticipated, one should avoid optoisolators with the base terminal of the phototransistor connected to a pin on the package, since the base terminal is sensitive to noise.

When a sufficently large potential difference is present between two adjacent conductors on a printed circuit board, a spark forms on the surface of the board between the conductors. The heat from the spark can form a blackened conductive path, called an "arc track." To prevent this degredation of the insulation one can specify a large distance between conductors. However, the dual-in-line package of common optoisolators has only 7.5 mm (0.3 inches) between input and output pins, so one can not conveniently increase the spacing beyond this distance. One can mill a slot in the printed circuit board, about 2 mm in width, between the input and output pins to obtain greater isolation (Motorola Optoelectronics Device Databook, 1983, p.4-4).

A particular disadvantage of optoisolators is that they have a finite lifetime. The brightness of the LED source decreases to about half of its initial level during a time of the order of 10^5 hours (10^5 hours = 11 years). The rate of degradation is greater at larger values of LED currents. This phenomenon is not widely discussed in the literature. One of the few references is a brief note by Lopez, Garcia, and Munoz (1977).

The LED in an optoisolator generally requires protection from being driven into reverse breakdown. Fig. 9-3a shows a standard cicuit that protects the LED from operation in reverse breakdown, since the 1N4447 diode will conduct at about -0.6 volts across the LED. The reverse breakdown voltage of the LED is typically between -5 and -20 volts, so the 1N4447 provides adequate protection.

The circuit in Fig. 9-3b provides comprehensive protection to the LED. When v_{IN} is positive, the avalanche diode protects the LED from operation with

(a)

(b)

IN4447

R

R₁

R₂

v_{in}
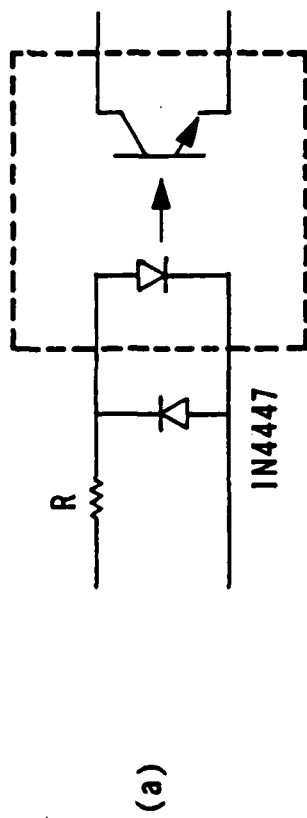
Fig.9-3
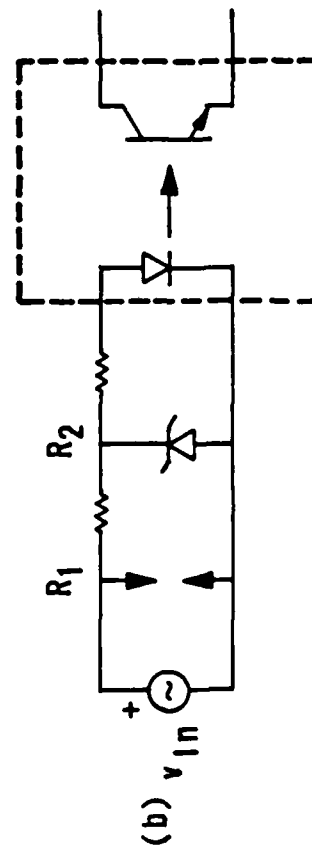
excessive current when the LED is forward-biased. The minimum value of R_2 is calculated from

$$R_2 \geq (V_Z - V_D)/\max(i_D)$$

where V_Z is the breakdown voltage of the avalanche diode, and V_D is the voltage drop across the LED when it is operating at its maximum rated curent, $\max(i_D)$. We could use the voltage drop at normal operating currents (e.g. 10 mA) for V_D in order to obtain a safety margin. In the absence of other design criteria, V_Z should be about 6.8 volts since zener or avalanche diodes have a minimum incremental impedance, $\Delta v/\Delta i$, at this value of V_Z. When V_{IN} is negative, the avalanche diode is forward-biased and protects the LED in the same way as the switching diode in Fig. 9-3a.

The value of R_1 can be obtained from

$$R_1 + R_2 = (V_{IN} - V_D)/i_D$$

where V_{IN} is the high level of the input source during <u>normal</u> (no overstress) operation of the circuit and (V_D, i_D) is the voltage across the LED and the current through the LED when the LED is on. We want R_1 to be as large as possible to protect the avalanche diode, so we use the minimum value of R_2. We remark that the value of R_1 can be much larger than the 20 Ω to 50 Ω that is commonly included between a spark gap and avalanche diode.

Example

Given: $\max(i_D) = 50$ mA $V_Z = 6.8$ volts

$V_D = 1.2$ volts at $i_D = 10$ mA

$V_{IN} = 5$ volts

Solution: $R_2 \geq 110$ Ω

$R_1 + R_2 = 380$ Ω

Let $R_2 = 110 \ \Omega$, then $R_1 = 270 \ \Omega$.

Notice that if the avalanche diode has a 5 watt steady-state rating, the maximum current in the reverse-biased avalanche diode, 735 mA, occurs when there is 205 volts across the spark gap. This is more than enough to fire a 90 or 150 volt spark gap, so this circuit is well-coordinated and will withstand continuous overstress of either polarity.

The optoisolator requires a power supply for the output side of the device. This is a serious disadvantage when the optoisolator is used as a transmitter, since it demands an independent power supply. An independent power supply is required since one dares not use the same power supply to transmit signals into an environment that contains transients and to operate vulnerable electronic devices, since the transient could be conducted directly to the vulnerable devices by the power supply.

Another possible disadvantge to optoisolators is that they are inherently uni-directional: information can only pass from the LED side to the photodetector side. Thus a pair of wires can not be used for bidirectional communications when optoisolators are connected at the end of the cable. This restriction could increase the number of wires (and the total mass) that are required for communication, a serious disadvantage aboard aircraft and missiles.

Unlike the isolation transformer, the optoisolator has an inherent ability to transfer continuous or DC signals. This is a particular advantage with relatively slow logic communications (e.g. 300 baud), since a transformer

would distort these signals owing to inadequate low-frequency response.

Optoisolators are usually reserved for digital interfaces owing to distortion of analog data with simple optoisolator circuits that have a switching device at the output port. There are a few optoisolators that are designed for linear operation in analog circuits. The General Electric H11F1 optoisolator has a photosensitive JFET at the output port. The value of V_{DS}/I_D (the output port resistance) is a linear function of the LED current. The Motorola MOC5010 optoisolator has a linear amplifier output for converting input current variations to output voltage variations.

PARASITIC INDUCTANCE

INTRODUCTION

When current passes through a conductor a magnetic field is created
(Ampere's Law). This implies that all conductors have self-inductance. When
the inductance is gratuitous (or not deliberately included), it is called
"parasitic inductance."

The amount of parasitic inductance is a critical issue when transients
are shunted to ground through non-linear components (e.g. spark gaps,
varistors, avalanche diodes, etc.). The voltage drop across the shunt path
consists of two terms: (1) the voltage due to the parasitic inductance, L
dI/dt, and (2) the voltage due to the current in the non-linear device. In
many practical situations the voltage due to the parasitic inductance is
greater than the voltage due to the current in the non-linear device. In fact
the inclusion of the non-linear device in the circuit may be vitiated by
excessive parasitic inductance. Sherwood (1977) says a "gas discharge device
with 0.5 inch leads on either side has lost a large percentage of its
effectiveness under fast rise time surge conditions."

We emphasize this point with an example. Suppose we have a 10 mm length
of 18 gauge copper wire (1 mm in diameter). This short piece of wire will
have an inductance of about 5 nH at 10 MHz. When dI/dt is 10^{11} A/s, which is
typical of lightning return strokes, 500 volts will be developed along this
"short-circuit" by the parasitic inductance.

Clark (1975) used a 100 A pulse with a 5 ns rise time (dI/dt = 2×10^{10}

A/s) to evaluate the effect of lead length on clamping voltage of avalanche diodes. He found the following values of voltage across an avalanche diode in a DO-13 metal case due to parasitic inductance:

7.6 cm lead length	1100 volts
3.2 cm lead length	700 volts
DO-13 case with no leads	100 volts

The diode case with no leads appears to have about 5 nH of parasitic inductance. Clark and Winters (1973, p.25, 50) gives a value of "approximately" 10 nH for the parasitic inductance of a DO-13 case with no leads.

Fisher (1978) discussed the effects of parasitic inductance in metal oxide varistors. This work emphasizes the importance of keeping the leads as short as possible.

Electronic circuit designers often work with currents of the order of 10 mA. Therefore, it is difficult for them to appreciate the large values of dI/dt that exist in transients, such as 10^{11} A/s during a lightning return stroke. In order to obtain a peak value of dI/dt equal to 10^{11} A/s from a sinusoidal waveform with a 10 mA peak current, we require a frequency of 1.6×10^{12} Hz. (This frequency has a free-space wavelength of about 0.2 mm.) Such a large frequency is way beyond the range of frequencies that are familiar to circuit designers, so they will not experience such values of dI/dt in their low-current work.

Obviously, one must minimize the length of conductors in shunt paths in order to minimize the parasitic inductance and maintain the good voltage

clamping properties of the non-linear device. What is not obvious is how to do this. For example, a diode in a DO-35 package has about 3 nH of parasitic inductance when its leads have zero length. In practice, we will have more inductance than 3 nH because we need a finite lead length.

A particularly insidious practice is to mount components on a printed circuit board. In order to avoid mechanical stress on the seal between the component lead and case, one must leave about 2 mm of straight lead before forming a bend. Each lead will have one right angle bend, so there will be a total of two right angle bends for each component case in the shunt path. These right angle bends are associated with increased parasitic inductance.

An even worse practice is to use components that have crimped leads to maintain proper spacing between the printed circuit board and the component. Crimped leads are commonly available for capacitors and varistors. The bends in each crimp, as well as the extra lead length, adds parasitic inductance.

While inductance is shunt elements is undesirable, inductance in series elements is generally desirable. The only exception to this rule would be when the signal frequency is greater than the highest significant frequency in the spectrum of the transient.

INDUCTANCE DOWNSTREAM

We can force the transient current to pass through a shunt element with parasitic inductance by placing a greater inductance downstream from the shunt element. In Fig. 10-1, L_1 is the parasitic inductance in a shunt element,
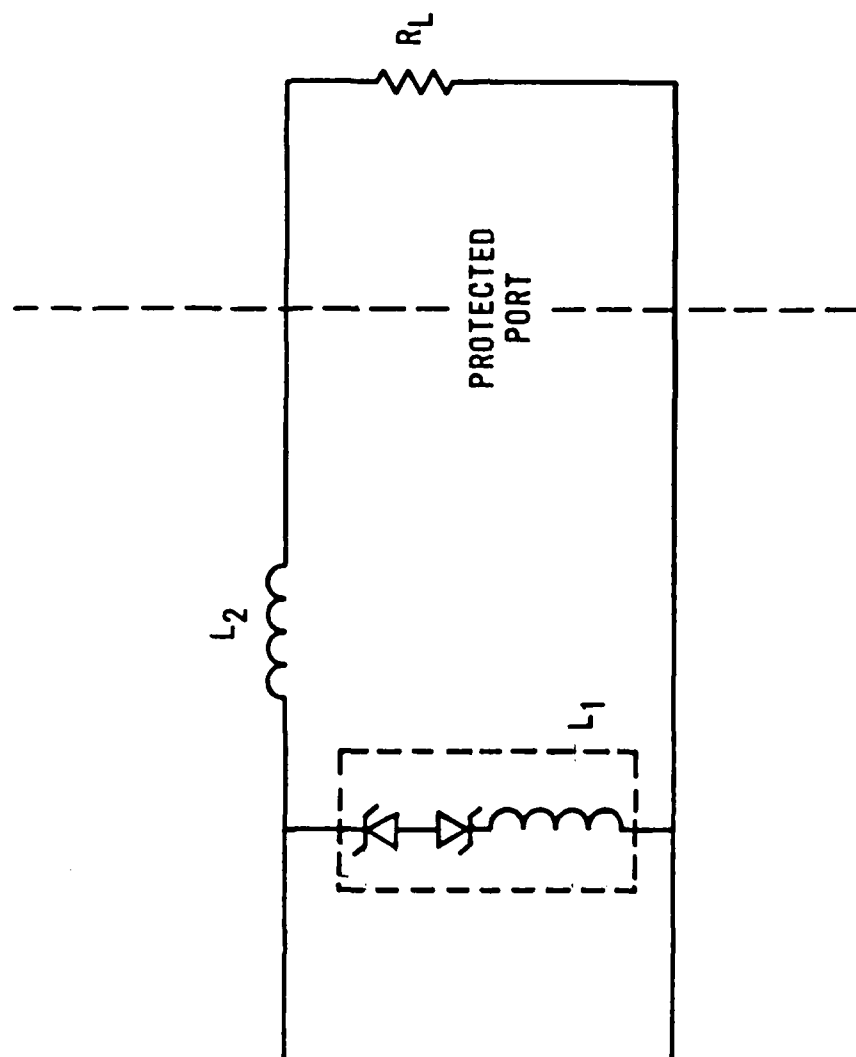
Fig. 10-1

which is shown as a bipolar avalanche diode, and L_2 is the inductance that is deliberately inserted between the shunt element and the protected port. If the protected load is not contiguous to the protect port, we must include the series inductance of connecting wires in the value of L_2. While L_2 is effective in forcing most of the transient to pass through the avalanche diodes rather than through the protected load, R_L, L_2 also has an undesirable effect. The avalanche diodes have a very small value of $\Delta v/\Delta i$ which clamps the transient voltage. By interposing L_2 between the diodes and the protected load, the voltage across the load is not as tightly clamped.

The inductance L_2 should not be a component. Rather, if it is used at all, it should be a trace on the printed circuit board that is about 2 or 3 times the distance through the shunt path that contains L_1 and has more right angle bends than the shunt path that contains L_1. In this way we are assured by construction that the parasitic inductance L_2 is greater than L_1. If the load to be protected is not within a few centimeters of the avalanche diode, the parasitic inductance in a straight conductor between the diode and the load will have sufficient inductance for L_2.

FOUR TERMINAL STRUCTURE

The parasitic inductance in a shunt path can be reduced to less than 1 nH by using a four terminal construction that was patented by Clark (1982) for avalanche diodes, see Fig. 10-2. Clark mounted the diodes on a strip of material as shown in Fig. 10-2a. A small piece of bent metal was used to connect the upper terminal of the diodes to the other strip, as shown in Fig. 10-2b. The four pins were then bent at a right angle to the plane of the

Fig.10-2
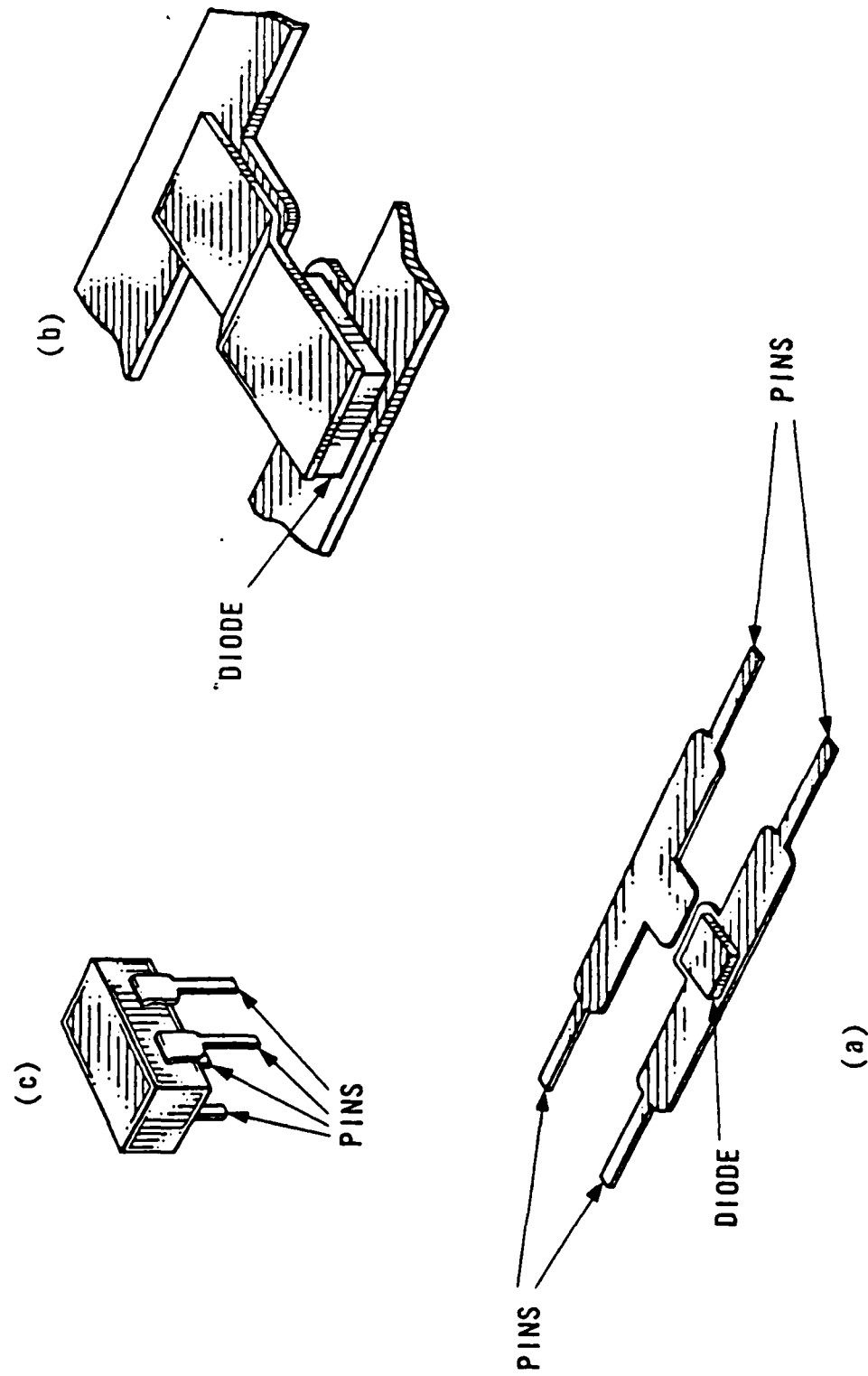
diode and two strips, and the assembly was molded in a plastic package as
shown in Fig. 10-2c. The inductance that is naturally present in any
conductor has been relocated in series with the transient (and signal) path by
the four terminal construction, as shown in Fig. 10-3. The four terminal
structure is commonly known as a "zero-inductance" package, but this is
inappropriate. The inductance is still present, but it has been moved to a
location where it does no harm.

Four terminal construction is also used in electrolytic capacitors for
use at frequencies between about 10 kHz and 1 MHz (Bowling, 1979). These four
terminal electrolytic capacitors are used as in switching power supplies.
They may be useful in filters for a DC voltage regulator that is located a
long distance from the transformer/rectifier module. However, the parallel
combination of an ordinary two-terminal electrolytic capacitor and two ceramic
capacitors will give a smaller impedance over a wider range of frequencies
than one four-terminal aluminum electrolytic, and at less cost.

Parasitic inductance of capacitors

If one plots the magnitude of impedance of a capacitor as a function of
frequency, f, one obtains a plot such as shown in Fig. 10-4. When the
impedance is proportional to $1/f$, the capacitor is behaving in the ideal way.
However, for all lumped element capacitors (the kind used in circuits below a
few hundred megahertz), there is a resonance frequency f_0 due to internal
inductance of the capacitor. At frequencies greater than f_0, the impedance of
the capacitor is dominated by the inductance term, and the impedance is then
proportional to f. You can estimate the internal inductance of the capacitor
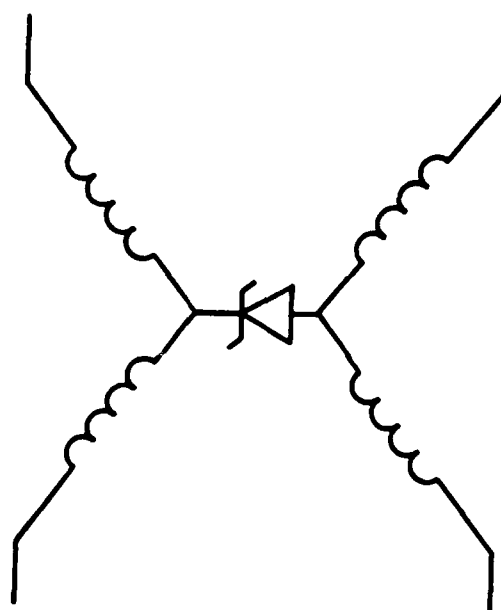
Fig.10-3
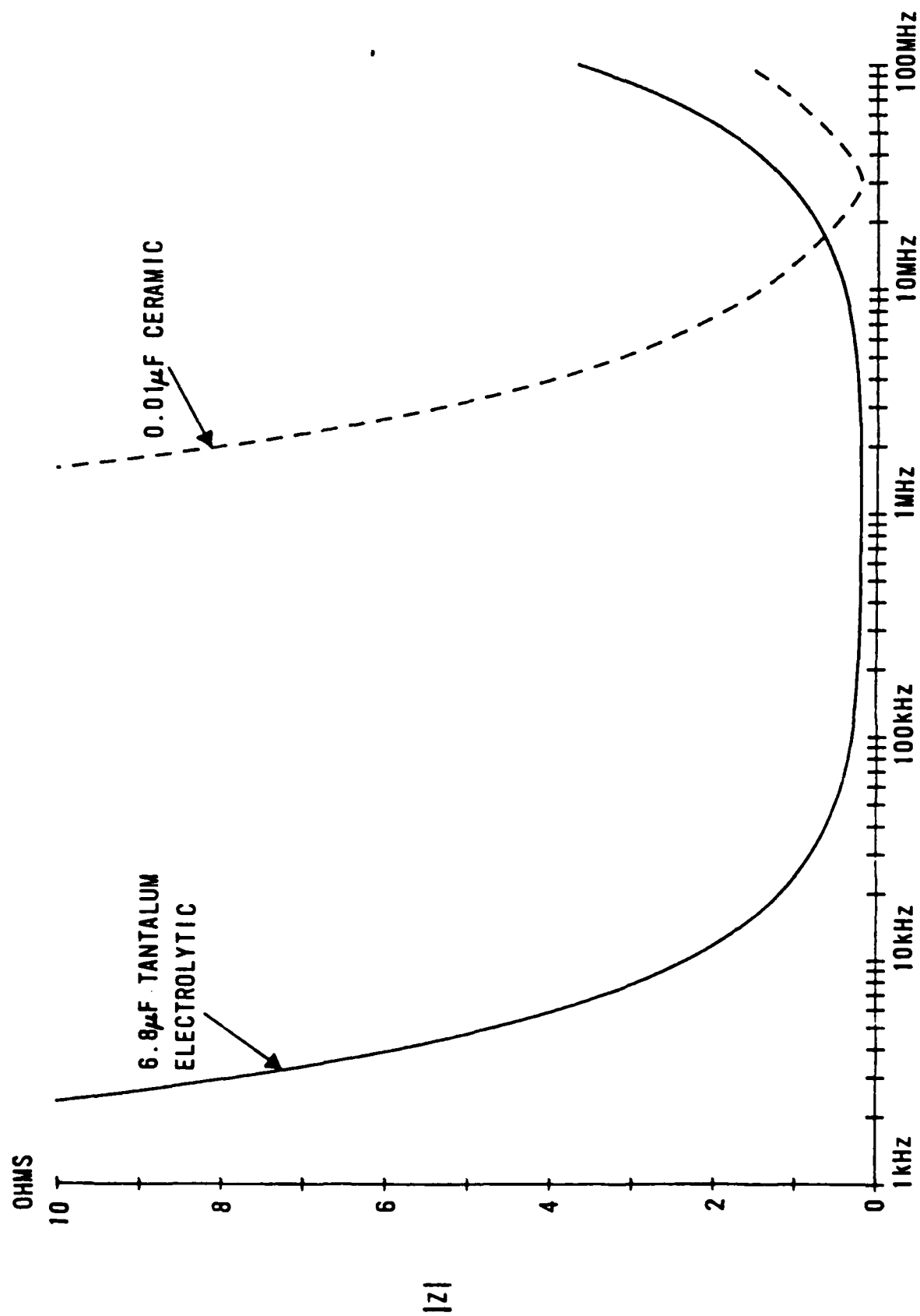
Fig.10-4

from knowledge of the resonance frequency and the capacitance value

$$L = 1/(\omega_0^2 C) \qquad \text{where } \omega_0 = 2 \pi f_0$$

Clearly, a real capacitor is not useful as a capacitor at frequencies that are much greater than f_0. Typical values of f_0 for non-polar lumped element capacitors are between 1 MHz and 100 MHz; typical values of parasitic inductance are between 1 nH and 50 nH. Woody (1983) measured values of parasitic inductance between 21 nH and 37 nH for 29 different capacitors.

Bypassing Power Supplies

Noise on power supply lines can cause unsatisfactory operation of the system during normal conditions. The manner in which noise appears on power supply lines, and how one removes it, are related to parasitic inductance. Therefore it is appropriate to treat this subject here, as well as in the chapter on DC power supplies. Noise is generated by varying currents in the wires that are used to connect the power supply to the load. The load is composed of the various amplifiers and other devices that require power from the supply. The wires that connect the power supply to the load have a small amount of resistance and inductance as shown in Fig. 10-5.

We will now analyze the effect of the inherent resistance, R, and inherent inductance, L, in the power supply connections. We will discuss a numerical example with values that are from a realistic situation. Suppose that each wire that connects a power supply to amplifier A_2 is a 50 centimetre length of copper wire with a diameter of 0.5 millimetre (known as 24-gauge wire). Then R will be 0.042 ohms. The inductance, L, is difficult to predict

but measurements show that it is of the order of 1 microhenry.

Suppose that the voltage source is sinusoidal:

$$V_{in} = 0.4 \sin(2\pi \, 10^5 t) \text{ volts}$$

The output current of amplifier A_2 is the output voltage, which has a 10 volt amplitude, divided by the parallel combination of 100 Ω and 10 kΩ, which is a sinusoidal waveform with an amplitude of 101 mA. This output current must come from the power supply lines. The total current drawn by amplifier A_2 from the power supply is approximately the quiescent current plus the output current. The quiescent current is the current in the power supply terminals when $V_{out} = 0$. Let us use 5 mA for the quiescent current. The total power supply current for amplifier A_2 is

$$5 \text{ mA} + 101 \text{ mA} \sin (2\pi \, 10^5 t)$$

Similar analysis tells us that amplifier A_1 has a total power supply current of

$$5 \text{ mA} - 1.2 \text{ mA} \sin(2\pi \, 10^5 t)$$

The total current at point X in Fig. 10-5 is about 10 mA + 100 mA sin(2π $10^5 t$). This current, I, causes a voltage drop between the +15 volt supply and point X of IR + L(dI/dt). The IR term has a peak value of about 4.6 millivolt. The inductive voltage drop has a peak value of 63 millivolts, much larger than the IR voltage drop. The peak of the IR and L(dI/dt) voltage drops are not in phase, since dI/dt is a maximum (or a minimum) when I is zero.

What is the value of the power supply noise at point Y, where the power supply is connected to amplifier A_1? Point Y will see all of the noise that we found above at point X, plus an additional amount due to the resistance,
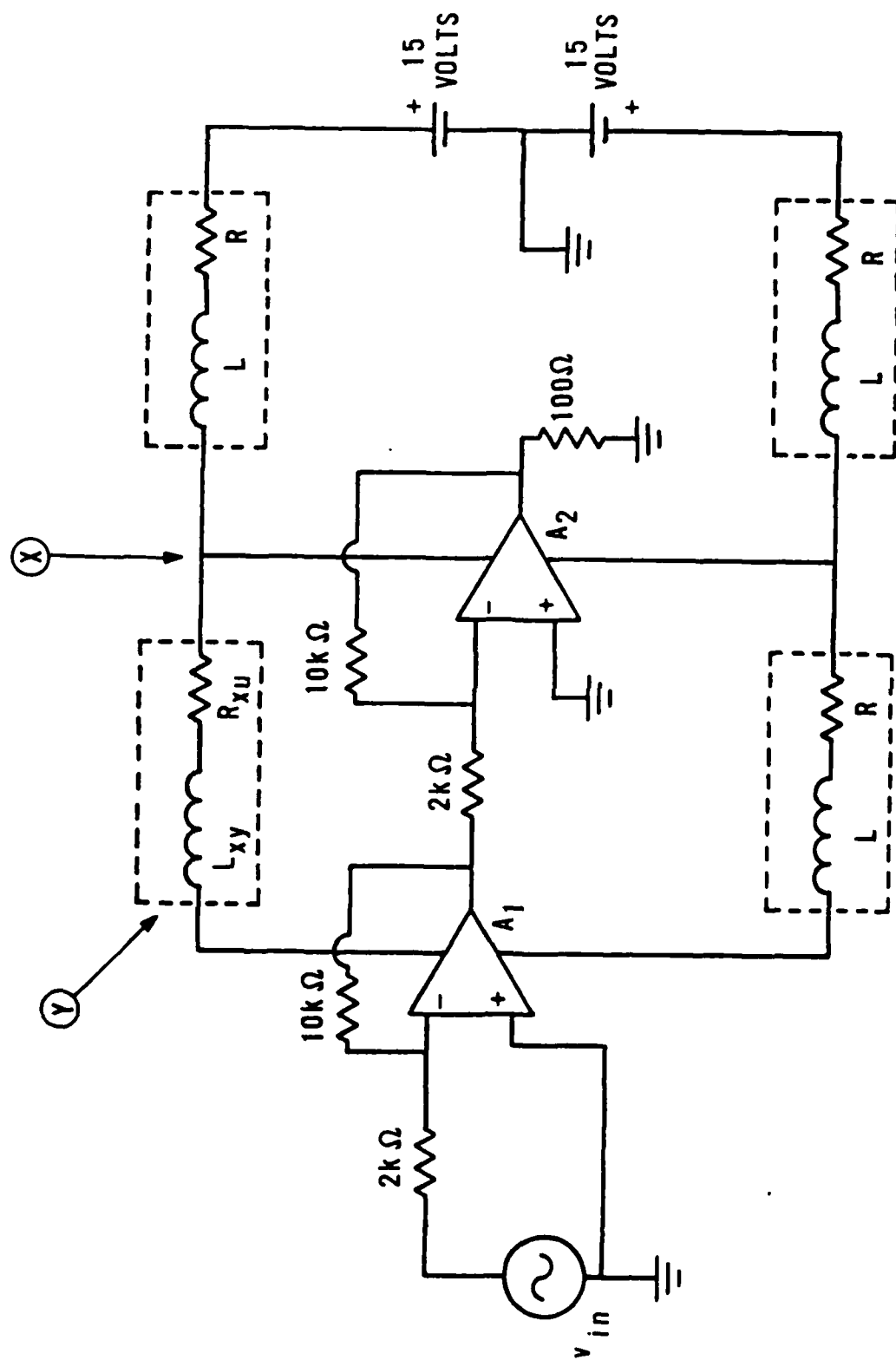
Fig.10-5

R_{xy}, and inductance, L_{xy}, between points X and Y. If the wire between X and Y is short, and the current to A_1 is small compared to the current to A_2, we can say that the noise on the power supply has approximately the same amplitude at points X and Y. Then most of the noise at amplifier A_1 (point Y in Fig. 10-5) is due to current drawn by amplifier A_2. This is an important lesson.

Amplifier A_2 can "talk" to amplifier A_1 via the power supply connections. Voltage fluctuations at the power supply terminals of an amplifier will result in some change in output voltage of that amplifier. Logic devices with edge-triggered inputs (e.g. edge-triggered flip-flops) are quite sensitive to noise. Noise on power supply lines during normal operation of the system can produce unsatisfactory operation of the system.

The standard cure for noise on the power supply lines is to use "bypass capacitors." Bypass capacitors are connected between each power supply terminal and ground. These capacitors must be located near the load as shown in Fig. 10-6. The function of bypass capacitors can be understood from either of two perspectives.

1. The parasitic series inductance in the power supply lines increases the output impedance of the power supply at high frequencies. The bypass capacitors provide a source of charge (or current) that has a much smaller output impedance, owing to the shorter distance (and therefore less parasitic inductance) between the capacitor and load.

2. We can think of an ideal DC power supply conductor as an incremental ground (because its voltage does not change with time). The parasitic inductance (and resistance) in the power supply wiring inserts an impedance between the ideal DC voltge source and the power supply

15 VOLTS

0.01 μF

6.8 μF

15 VOLTS

0.01 μF

6.8 μF
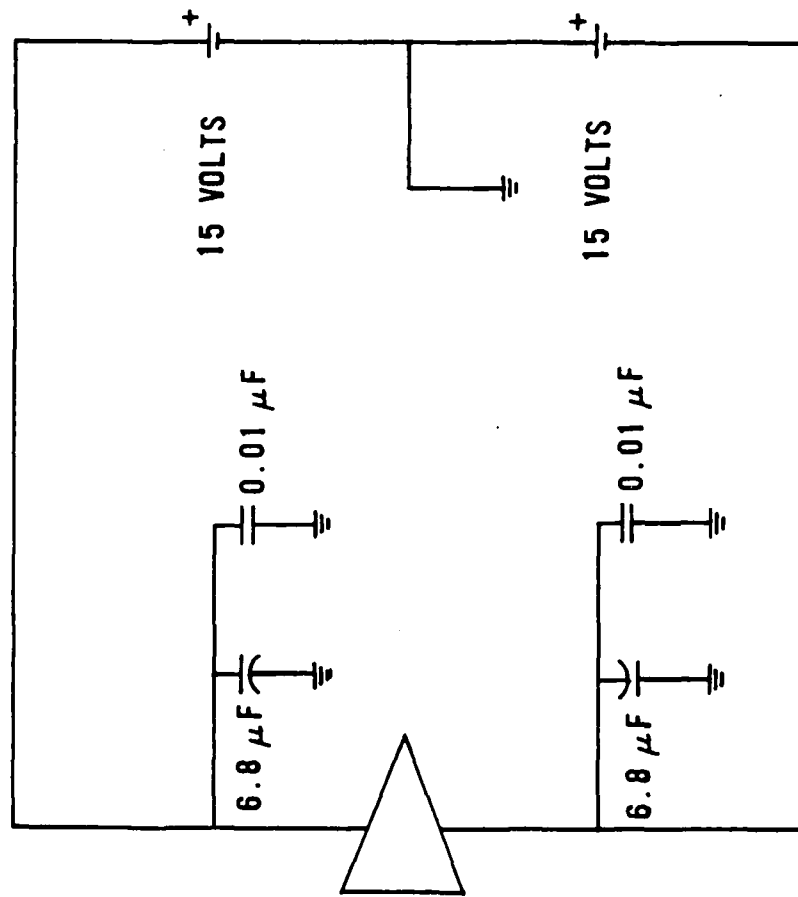
Fig.10-6

terminals of the load. Using bypass cpacitors at the power supply
terminals of the load provides a low impedance path to ground, and
removes voltage fluctuations at that point. In this view, power supply
bypass capacitors have the same function as the capacitor across the
emitter or source resistor in the self-bias common-emitter or common-
collector transistor amplifiers.

Because real capacitors have parasitic inductance and a resonance
frequency, it is standard practice to place several different kinds of
capacitors in parallel to form the bypass element. A typical selection would
be a tantalum electrolytic capacitor (1 µF to 10 µF) in parallel with one or
more ceramic capacitors (0.01 µF to 0.1 µF). The exact value of the bypass
capacitance is not critical; it is often difficult to see the effect of a
factor of two change in capacitance. What is critical is that the bypass
capacitors maintain a suitably low impedance at all frequencies of interest.
Fig. 10-7 shows the magnitude of reactance vs. frequency for the parallel
combination of three capacitors with the following properties:

	C	L	series resistance
tantalum electrolytic	6.8 µF	5.8 nH	0.2 Ω
ceramic disc	0.1 µF	7.0 nH	0.2 Ω
ceramic CK06 style	0.01 µF	2.8 nH	0.2 Ω

The 0.1 µF ceramic disk capacitor is not essential, but it does reduce the
impedance between 10 MHz and 20 MHz by about a factor of two. At frequencies
below about 30 kHz, the impedance of this capacitor network is not necessarily
small. However, the parasitic inductance of the power supply lines is usually
not a problem at these low frequencies.

Fig.10-7

100μF ALUMINUM ELECTROLYTIC
6.8μF TANTALUM
0.01μF AND 0.1μF CERAMIC
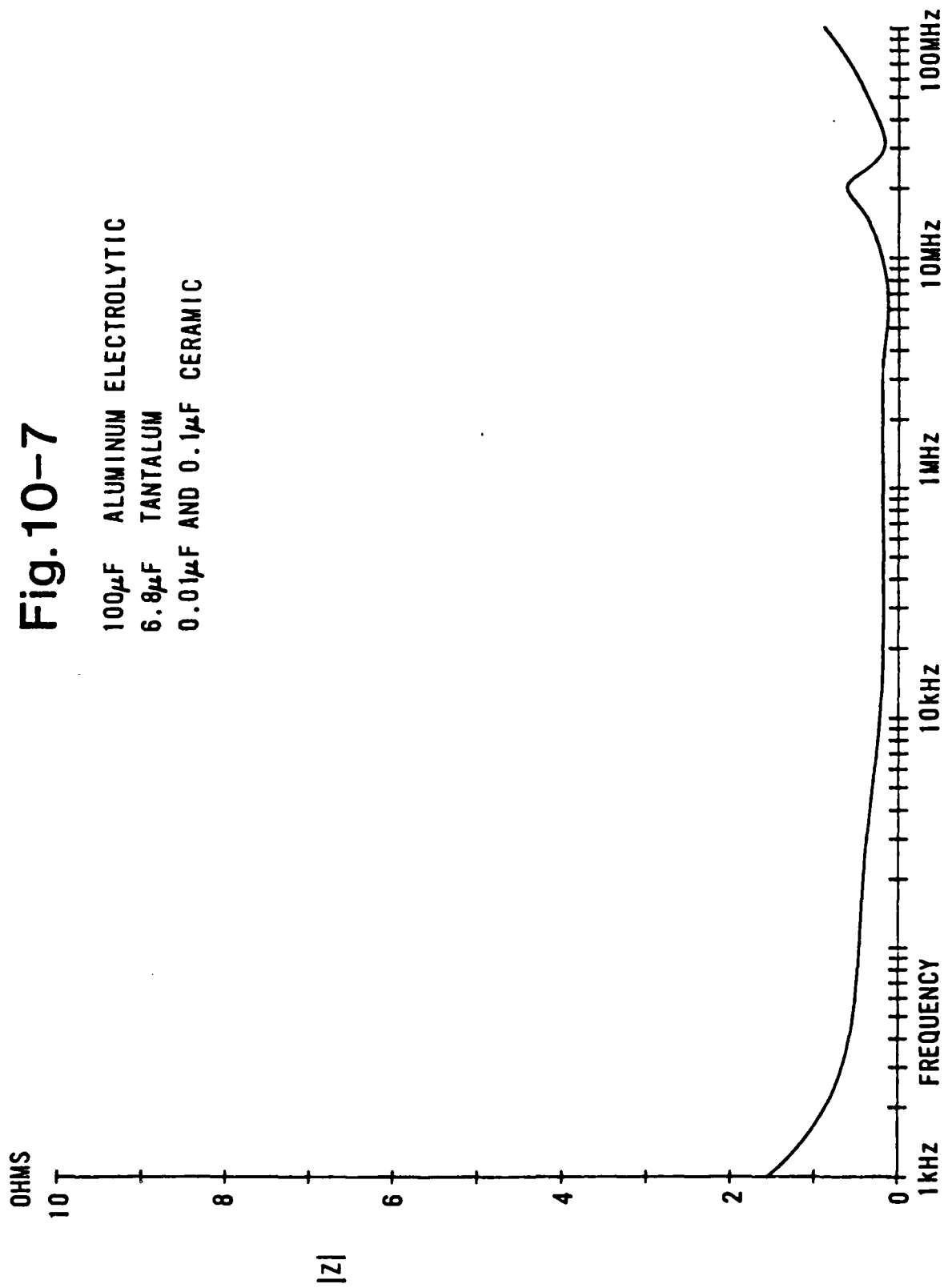
TRANSIENT PROTECTION OF SIGNAL CIRCUITS

INTRODUCTION

In this chapter we shall be concerned with applications of transient protection to signal circuits. "Signal circuits" are defined here as those in which the magnitude of the voltage is less than about 15 volts, and the magnitude of the current is less than about 50 mA. The bandwidth of a signal circuit may extend from DC to more than 1 MHz.

Line receivers have their inputs connected to a signal line and they receive data from a distant source. Line drivers have their outputs connected to a signal line and they transmit data to a distant receiver. Line receivers and line drivers are examples of "interface modules." Most of the transient protection problems on signal lines arise during protection of interface modules in a system, since the overstresses that are conducted on long cables from the exterior of a system tend to be much more severe than internally generated overstresses.

We shall first discuss some basic circuits that can be used on almost any signal line, regardless of whether it is analog or digital. Then we shall discuss some specific examples for analog circuits, followed by specific examples for digital circuits. The discussion of digital circuits will emphasize computer data lines.

SPARK GAP AND AVALANCHE DIODE CIRCUIT

Bodle and Hays (1957) invented the basic circuit that combines a spark

gap and avalanche diodes, as shown in Fig. 11-1, to provide comprehensive
protection from electrical overstresses. This circuit has been described by
many other authors, including Popp (1969), Greenwood (1971, p.318), Knox
(1973), and Clark (1976), among others.

The load is protected by an avalanche diode absolute value clipping
circuit, D_1 and D_2, which will be discussed in detail later. The resistor,
R_1, provides a large voltage across the spark gap when the current in the
avalanche diodes is large. If the voltage across the spark gap is
sufficiently large, the spark gap will conduct and shunt current away from the
avalanche diodes. The avalanche diodes protect the load; the spark gap and
resistor protect the avalanche diode

When the spark gap is conducting it acts in two ways to protect the
circuit. First, it dissipates some of the transient's energy as heat. Second,
its low impedance at the end of a transmission line causes some of the
transient to be reflected. The reflection phenomenon is useful for delaying
the dissipation of transients with large values of dI/dt. If we have 500
metres of line that has a short at each end, due to conducting spark gaps, the
round trip time delay is about 5 μs. After multiple reflections, of course,
the transient's energy must be dissipated in the protection circuit and the
resistance of the line.

A spark gap will be non-conductive until the voltage across the gap
exceeds the DC firing potential. The spark gap then operates in the glow
region for currents between about 1 and 100 mA with about 80 to 100 volts
across the gap. As the current increases above 100 to 500 mA, the spark gap

OVERSTRESS

R₁

L
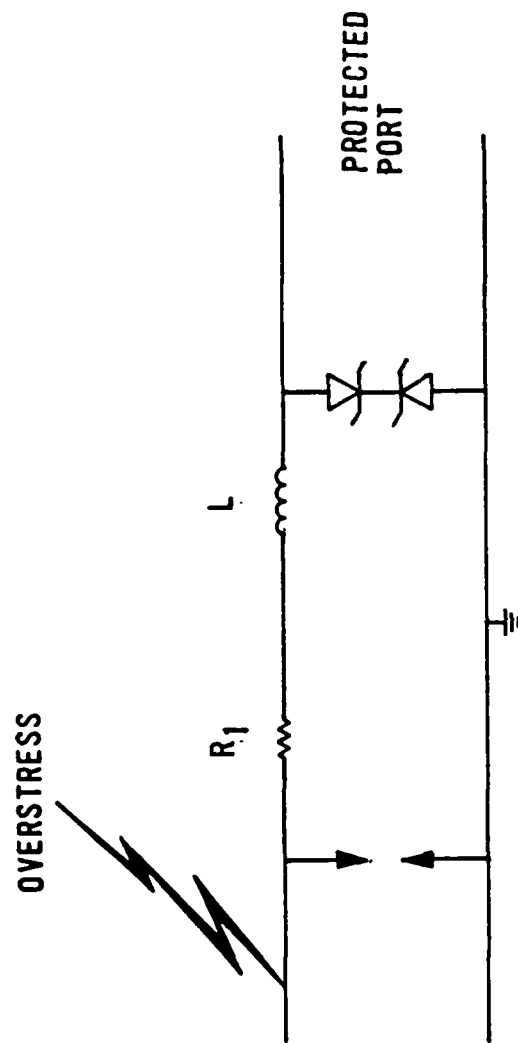
PROTECTED
PORT

Fig.11-1

will operate in the arc region with 10 to 20 volts across the gap.

The DC firing voltage is not the only important specification in transient protection, owing to the time delay between the application of a transient voltage and the onset of appreciable conduction in the spark gap. The value of the time delay is complicated to predict and is a function of the rate of rise of the potential as well as the spark gap parameters. For example, if 300 volts is applied with a rise time of less than 0.05 μs to a spark gap with a 150 volt DC firing voltage, the gap will typically remain non-conducting for about 0.5 to 1.0 μs, then operate in the glow region for about 0.5 to 1.5 μs, and then operate in the arc region. The point to be made is that a spark gap can be (briefly) non-conducting with a potential across it that is several times its DC firing potential. Furthermore, this high potential can remain across the spark gap for durations on the order of a microsecond. This is long enough to damage vulnerable electronic circuits if other protective devices (e.g. avalanche diodes) are not used. The main reason for including spark gaps in transient protection circuits is that they excel at shunting currents of the order of 5 to 10 kiloamperes away from vulnerable circuits.

It should be mentioned that if the spark gap conducts when the circuit is connected across a continuous voltage source of at least 20 volts with the ability to source at least 100 mA, serious damage could result, owing to follow-current in the spark gap. Unless a fuse or circuit breaker interrupts the process, either the spark gap, the printed circuit board, and/or the connecting cable will be permanently damaged. If such situations are anticipated, a current-limiting device must be included between the line and

the spark gap, as discussed in Chapter 8. Possible current-limiting devices
include a circuit breaker, a positive temperature coefficient (PTC) resistor,
a resistor with a large steady-state power rating, or a fuse. Also the DC
firing voltage of the spark gap should be increased to a value between 200 V
and 300 V.

The avalanche diode voltage rating is specified so that it will not
conduct during normal operation of the system. This implies that the
avalanche diode should have a large breakdown voltage. However, there are two
reasons why the diode should have the smallest possible breakdown voltage.
First, small values of breakdown voltage allow the protection circuit to clamp
at a lower level and thus provide less stress on the devices to be protected.
Second, avalanche diodes can be destroyed by large amounts of power or energy
during overstresses. Diodes with smaller values of breakdown voltage can
tolerate larger surge currents, since power or energy are proportional to both
breakdown voltage and current.

The avalanche diode breakdown voltage will usually not be more than a few
volts greater than the supply voltage for the active circuit that transmits
the signals. In some situations the avalanche diode breakdown voltage can be
less than the supply voltage for the active circuit that transmits the signal.
The effects of tolerance should be considered when specifying the avalanche
diode breakdown voltage. There is often a very small cost increase (e.g $0.10
in a item with a cost of $1.5) associated with diodes with a 5% tolerance,
when compared to diodes with a 10% or 20% tolerance. In many situations the
smaller tolerance devices will be worth the extra cost, since the mean
clamping level can be smaller without interfering with the normal operation of

the system.

Transient suppression diodes are available in "bipolar" models, for which D_1 and D_2 are combined into a single package. This reduces the cost of components and circuit assembly time. Moreover, the bipolar diodes permit a shorter distance on the circuit board compared to two unipolar diodes that are connected in series. This shorter distance will reduce the parasitic inductance that is in series with the shunt path through the bipolar avalanche diodes.

It is generally recommended that the avalanche diodes in Fig. 11-1 have a steady-state power dissipation rating between 1 W and 5 W. However, if the value of R_1 can be made sufficiently large, the avalanche diode could have a smaller power rating, such as 0.5 W. The only advantage to using diodes with small power ratings is a reduction in mass and volume of the circuit. Diodes with a larger power rating, and particularly transient suppression diodes that are designed for very large instanteous power ratings, offer increased assurance that the protection circuit will survive.

The series resistor, R_1, limits the current through the avalanche diodes, D_1 and D_2, to a safe value. The resistor should be physically large to prevent "flash-over". Suitable types include 2 watt carbon composition and wirewound resistors with a steady-state power rating of at least 2 watts. Metal film and carbon film resistors are not suitable for limiting large current pulses.

The resistance value of resistor R_1 is calculated by the following

conservative procedure and Eqn. 1.

$$(V_G - V_z) \; V_z \; / \; P = R_1 \qquad\qquad (1)$$

P is the maximum steady-state avalanche diode power rating specified by the manufacturer. V_z is the avalanche diode reverse breakdown voltage. V_G is the maximum value of the DC firing voltage of the spark gap. I recommend increasing the value of V_z in Eqn. 1 by 20% from the nominal avalanche voltage specified by the manufacturer. This compensates for the increase in magnitude of the breakdown voltage at large currents (due to the incremental resistance of the diode and the temperature coefficient) and for the 5% tolerances of the resistor and avalanche breakdown voltage.

During transients with rapid risetimes the spark gap will conduct at a voltage than can be several times greater than V_G. During the microseconds that the spark gap is non-conducting, the current in the avalanche diode will be greater than its steady-state maximum value. This will not damage the avalanche diode because the overload time is so brief.

Using Eqn. 1 to determe the value of resistance R_1 is perhaps too conservative. A constant voltage source of magnitude V_G, which is typically between 90 and 150 volts DC, is very uncommon in modern industrial environments. Therefore it is unlikely that the transient protection circuit would be connected for an indefinite time to such a large DC voltage. It is however, conceivable that the circuit could accidentally be connected to a sinusoidal voltage source (e.g. 120 volt rms line). We can make the simplifying assumption that the avalanche diode has a constant voltage, V_z,

across it when it is reverse biased and conducting. We also assume that V_G is much greater than V_Z so that each avalanche diode is reverse-biased and conducting during essentially half of the period of the sinusoidal voltage. We then obtain Eqn. 2 for the minimum value of R_1.

$$R_1 = (\ (V_G/\pi) - V_Z \ ) \ (V_Z/P) \qquad\qquad (2)$$

If both avalanche diodes are in the same package, then V_G is divided by $\pi/2$, not π, in Eqn. 2.

The conventional method for calculating the value of R_1 uses a standard model for the transient overstress. Huddleston and Bush (1975) advocated the following method. Let

V_S = maximum expected transient voltage (e.g. 10×1000 μs waveform)

I_S = manufacturer's rated peak surge current of diode for same waveform as used for V_S

min(V_Z) = nominal avalanche diode breakdown voltage multiplied by worst-case factor for tolerance (e.g. by 0.95 for a ±5 % tolerance device)

Then use Eqn. 3 to determine the minimum value of R_1.

$$\min(R_1) = [V_S - \min(V_Z)]/ \ I_S \qquad\qquad (3)$$

The value of min(R_1) should be multiplied by a factor for tolerance to obtain the nominal value, e.g. multiply by 1.05 to convert min(R_1) to R_1 for a 5 % tolerance resistor. The value of R_1 that is obtained with the methods of Eqns. 1, 2 or 3 will probably be a non-standard value (e.g. 19 Ω instead of 20

Ω), so it should be increased to the next larger standard stock value.

Resistor R_1 forms a voltage divider with the load resistance. The value of R_1 should be as small to avoid excessive attenuation of the signal. On the transmitting end of the interface, resistor R_1 forms a low-pass filter with the line capacitance between signal and common conductors. When this circuit is used on the transmitting end of a long cable and high speed communications are required, the value of R_1 should be as small as possible. On the other hand, R_1 should be large in order to limit the current in the avalanche diodes. Clearly there must be a compromise for the value of R_1.

Many transient protection circuits adopt a value between about 5 Ω and 50 Ω for R_1. Certainly, if a larger value of R_1 is tolerable in a particular application, it should be used because increasing the value of R_1 increases the amount of transient protection in an economical way.

There are two particular ways out of the dilemma posed by requirements for both a large value of R_1 and a small value of R_1. One way is to use inductance to increase the series impedance between the spark gap and avalanche diodes. The other method is to use a positive temperature coefficient resistor.

Inductance can be inserted by using a wirewound resistor for R_1, and/or by placing ferrite bead(s) on the resistor leads. Typical values of inductance in this circuit are between 1 μH and 10 μH. Using ferrite beads is a particularly good idea if the highest frequency component in the signal has a frequency less than about 100 kHz. At higher frequencies, ferrite beads

become quite lossy and will provide significant attentuation of transients. Series inductance will distort a signal that has high frequency components, and inductance should be avoided in this situation.

Using inductance to increase the series impedance has the disadvantage of offering no protection from sustained overvoltages. However, positive temperature coefficient (PTC) devices, which were described in Chapter 8, do offer substantial protection from overvoltages which have durations of 0.1 second or longer.

Because the transients can have high voltages, the circuit designer should take precautions to avoid an arc path that could shunt resistor R_1. To intercept flash-over, should it occur, a wide ground band should be etched on both sides of the printed circuit board underneath resistor R_1.

The typical power line ground does not have minimal inductance due to its relatively long path and bends in conduit or cable. It would preferable to connect a separate conductor directly to a good local ground (e.g. copper water pipe) for any installation where peak currents greater than about 10 amperes were anticipated. Such a ground conductor should be flat copper braid with a width of at least 5 mm and connected directly to the common terminals of all of the spark gaps. For situations in which peak transient currents of less than about 10 amperes are anticipated, one can use the ground conductor in the signal cable.

DIODE CLAMPS TO POWER SUPPLY

The other basic transient protection circuit is to use diodes to clamp the signal line to the DC power supply line, as shown in Fig. 11-2. The device to be protected in Fig. 11-2 is U_1, which might be an amplifier, logic device, or other integrated circuit. The two DC power supplies for the U_1 are V_{CC} and V_{EE}. Each of these two supplies has a typical magnitude of 15 volts. Diode D_1 prevents the input terminal of U_1 from having a potential greater than V_{CC} + 0.7 volts. Similarly, diode D_2 prevents the input terminal of U_1 from having a potential less than $-(V_{EE}$ + 0.7 volts). Since most integrated circuits can continually withstand any potential at the input terminal that is between V_{CC} and $-V_{EE}$, this circuit provides good protection for the input terminal of U_1. The circuit of Fig. 11-2 must not be used if the input terminal can not withstand application of the power supply voltages.

It is common to use core-switching diodes, which were discussed in Chapter 6, for D_1 and D_2. These diodes are normally reverse-biased in the circuit of Fig. 11-2, and their leakage current, which is typically less than a picoampere, would not be expected to cause problems with U_1 in most applications. The shunt capacitance of D_1 and D_2 could couple noise from the power supply to the input of U_1. However, the capacitance of core-switching diodes is only a few picofarads. If the input impedance of U_1 is much smaller than the impedance of the diode's capacitance, the voltage divider that is formed by these two impedances will substantially attenuate the noise injected from the power supply. Power supply bypassing is necessary for other reasons and this bypassing will attenuate noise on the power supply lines.

We mention that the small parasitic shunt capacitance of this protection method makes it suitable for protecting high-frequency circuits. Of the
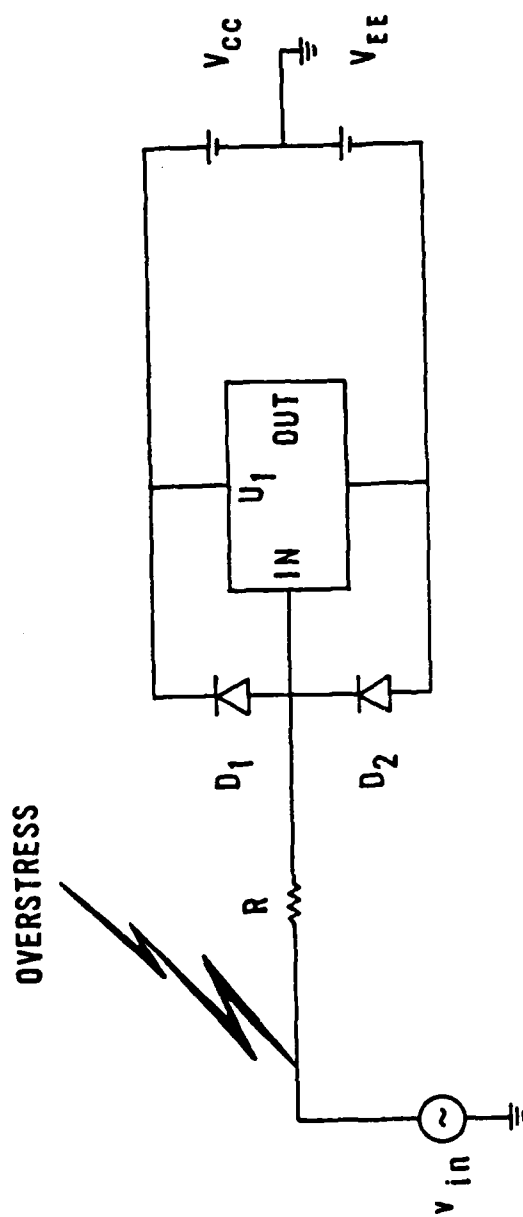
Fig.11-2

available transient protection components, only spark gaps and core-switching diodes have a parasitic capacitance of the order of picofarads. All other components have a larger capacitance.

The circuit in Fig. 11-2 has a serious problem. The overstress that threatened the input of U_1 was shunted to one of the power supply lines. It is possible that the overstress could damage U_1 or other devices that are connected to these DC power supply lines by injecting the overstress into the power supply terminals. If the overstress has a very small charge transport, e.g. less than 0.1 µC, then it can be shunted to ground through bypass capacitors without causing any distress, provided that the bypass capacitors are located near U_1. However, overstresses with a charge transport of substantially more than 0.1 µC should not be diverted with the circuit of Fig. 11-2. The value 0.1 µC comes from an arbitrary decision to limit power supply transients to less than 0.1 volts and use a 1 µF bypass capacitance. We emphasize that the circuit in Fig. 11-2 offers excellent protection of the input terminals from electrostatic discharge, a transient with a small charge transfer.

The value of the resistor R in Fig. 11-2 limits the current in diodes D_1 and D_2. It would be desirable to make the value of R no larger than 1 % of the input impedance of U_1, so as to avoid attenuation of the input signal. Core-switching diodes can tolerate transient currents of the order of 4 A for 1 µs, which is a 4 µC charge transfer. Because of concern about injecting noise on the power supply lines, we restricted the circuit in Fig. 11-2 to situations in which the charge transfer was less than 0.1 µC. The core-switching diodes are robust enough to survive these small transients without a

series resistance.

The general circuit concept of Fig. 11-2 is used to protect complementary metal oxide semiconductor (CMOS) logic integrated circuits. The input circuit of a basic CMOS cell is shown in Fig. 11-3, along with the transient protection circuit. What is particularly interesting in Fig. 11-3 is that the transient protection circuit is fabricated as part of the integrated circuit. The circuit has been described by Cergel (1974) and Pujol (1977). The resistor, R, has a value between 100 Ω and 2500 Ω, depending on manufacturer and the particular model of CMOS. Diode D_2 is distributed along the length of the series resistor R. The reverse breakdown voltage of the protection diodes is greater than the maximum value of $(V_{DD} - V_{SS})$, so the diodes will not operate in reverse-breakdown during sustained transients. The maximum steady-state input terminal current in a CMOS integrated circuit is 10 mA.

If the CMOS circuit is to be connected to a voltage source that has a level greater than V_{DD} or less than $-V_{SS}$, an external series resistor should be connected to the input terminal to reduce the current in the protection diodes to a safe level. If the external resistor has a value of 1 MΩ, the CMOS device can be connected to a source of 1 kV without stressing the protection diodes. The external resistance and input capacitance form a low-pass filter. The typical CMOS input capacitance is about 5 pF. With a 1 MΩ external resistance, we obtain a 5 μs time-constant or a 32 kHz bandwidth. This is much slower than the time constant of a few nanoseconds that is produced by the internal resistor R and the input capacitance. Unfortunately, nearly all robust transient protection circuits reduce the signal bandwidth.
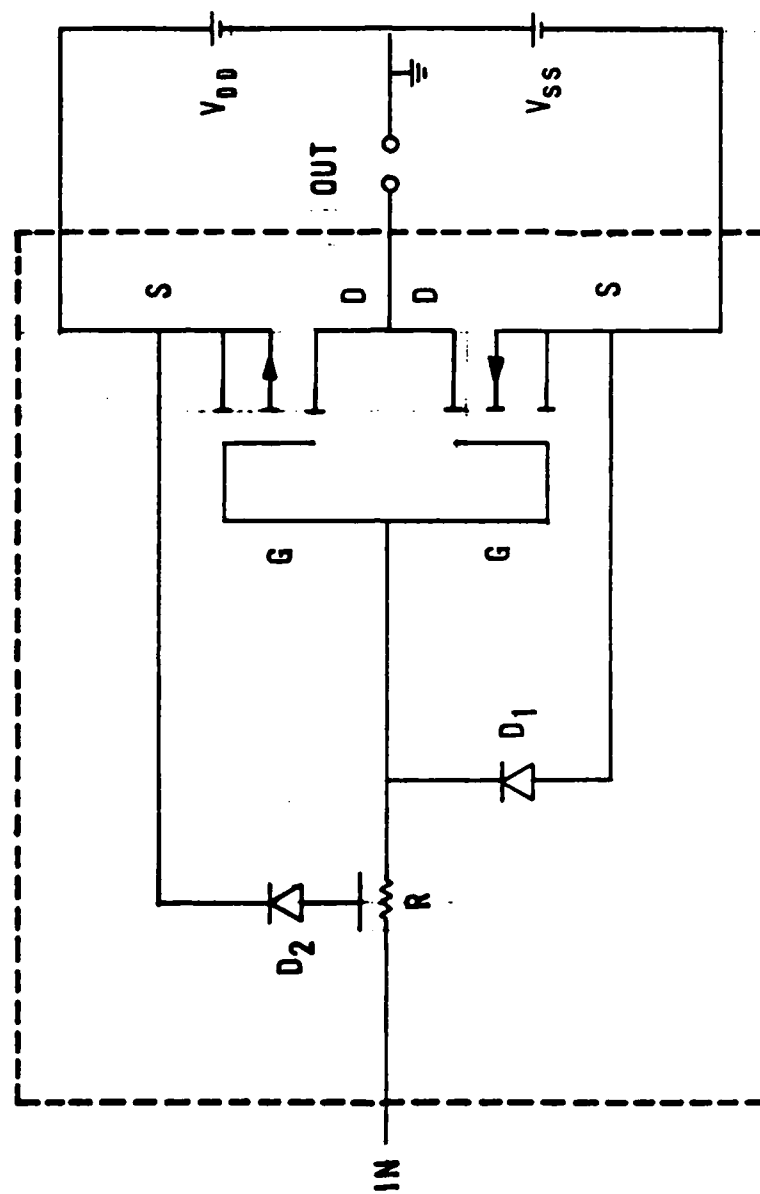
Fig.11-3

BALANCED LINE APPLICATIONS

Many long signal lines are balanced in order to reject noise, which often appears as a common-mode signal. These balanced signal lines require special consideration for transient protection. A basic spark gap and avalanche diode circuit for a balanced transmission line is shown in Fig. 11-4.

The spark gap must be a three electrode model, which was described in Chapter 3. Use of a pair of two electrode spark gaps can pass a large differential-mode voltage downstream if one, but not the other, of the two electrode gaps conducts.

The same value of R and L must be present in each side of the circuit in order to preserve the balanced configuration. The value of R can be determined in the same way as for the non-balanced circuit, which was presented in Fig. 11-1, by considering one-half of the balanced circuit and ignoring D_5 and D_6. In Eqns. 1 through 3 above, V_Z becomes the voltage for the series combination of D_1 and D_2; V_G is the DC firing voltage from either line to ground.

The inclusion of inductance L is particularly desirable in the balanced circuit. All of the commercially available three electrode spark gaps have DC firing voltages of at least 300 volts, because they are most commonly used for protecting telephone lines. This DC firing voltage is considerably larger than the 90 or 150 volt models that are commonly used on non-balanced signal lines. By placing series inductance between the 300 V DC spark gap and the avalanche diodes, we help create a large voltage across the spark gap during
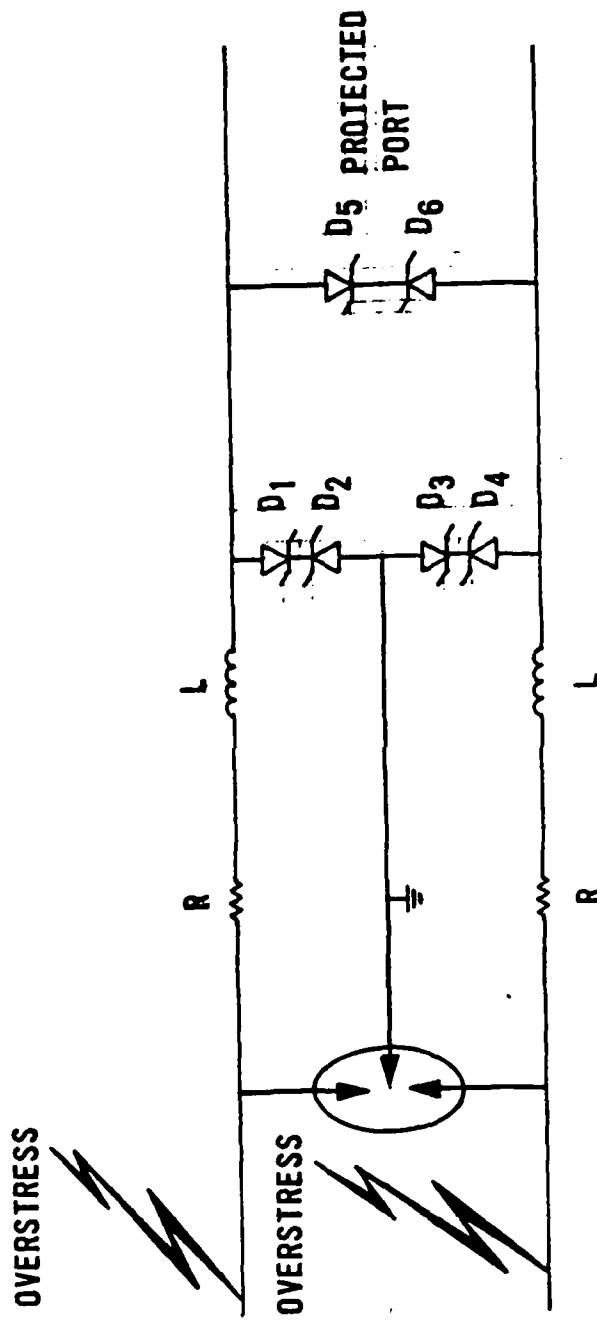
Fig.11-4

the initial part of a transient without increasing the series impedance during normal operation to an unacceptably large value. If the signal bandwidth extends to relatively high frequencies, the pair of inductors should be a bifilar choke, which was described in Chapter 8. A bifilar choke inserts a series inductance for common-mode signals, but has a negligibly small inductance for differential-mode signals.

Avalanche diodes D_1 and D_2 clamp the upper signal conductor in Fig. 11-4 to a potential between $\pm V_Z$ of ground. Diodes D_3 and D_4 clamp the lower signal conductor to a potential between $\pm V_Z$ of ground. If diodes D_5 and D_6 are absent, the maximum differential-mode output voltage could be as large as twice V_Z. This situation occurs, for example, when the upper signal conductor is at a potential $+V_Z$ from ground and the lower conductor is at a potential $-V_Z$ from ground. Inclusion of diodes D_5 and D_6 can limit the maximum differential-mode output voltage. In most cases, diodes D_1, D_2, D_3, D_4, D_5, and D_6 will have identical specifications, including breakdown voltage, V_Z.

During normal operation of a balanced line the upper signal conductor will be a potential of $V_d/2$ with respect to ground and the lower signal conductor will be at a potential of $-V_d/2$ with respect to ground. The value of V_d may be either positive or negative. In this way there is no common-mode voltage present during normal operation. This requires that the line drivers for a balanced line have two power supplies, one positive and one negative.

The RS-422 specification for computer data exchange on a balanced line does not require a zero common-mode voltage (although it is acceptable to use one). Most integrated circuit line drivers that conform to RS-422

specifications use only a single supply, which is usually +5 volts. The
voltage between one signal conductor and ground is typically either 0.2 or 3
volts, depending on whether the line is at the low or high state. This
implies that there will be a DC common-mode signal of 1.6 volts, which is
about 60 % of the differential-mode signal. This use of a single power supply
decreases the cost of the line driver system, since including a negative power
supply would substantially increase the cost owing to an extra secondary
winding on the power transformer, extra filter capacitor, and an extra voltage
regulator. The presence of a DC common-mode signal does not preclude the use
of a bifilar choke.

Smithson (1977) developed a protection circuit for a balanced computer
data line that is shown in Fig. 11-5. His circuit is interesting because he
used three stages of protection: spark gaps, metal oxide varistors, and zener
diodes. The use of varistors is significant. The varistors decreases the
stress on the diodes when the transient does not cause the spark gap to
conduct, in fact all transients with a peak of less than 300 volts will not
fire the spark gap. Moreover, the varistor also decreases the stress on the
diodes during the initial portion of transients before the spark gap fires.

It is interesting to estimate the varistor current and voltage when the
voltage across the spark gap is just equal to its 300 V DC firing voltage.
This condition occurs when there is about 30 A in the varistor and about 53
volts across the varistor. (This information is from the "maximum clamping
voltage" graphs in 1978 edition of the General Electric company varistor data
sheets.) When 30 A passes through the 8 Ω resistor between the spark gap and
varistor, 240 volts will be dropped across the resistor. The sum of these two

OVERSTRESS

OVERSTRESS

27 Ω

27 Ω

8 Ω

8 Ω

TO LOAD

4.7V
ZENERS

24 V
VARISTORS

300 VDC
SPARK
GAP
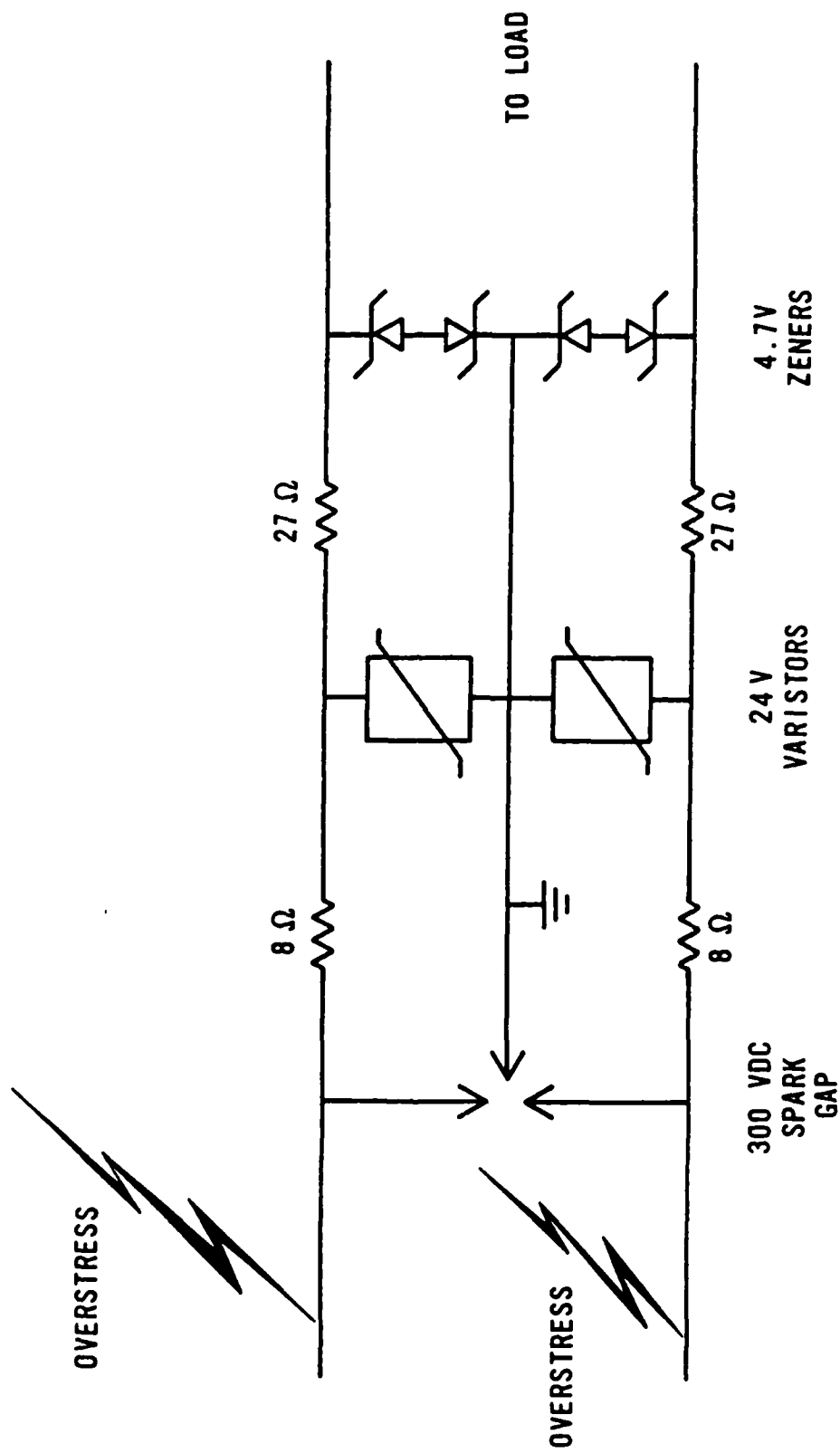
Fig.11-5

BALANCED LINE PROTECTOR

voltage drops is 293 volts, essentially the DC firing voltage of the spark
gap. That about 80% of the voltage across the spark gap comes from the drop
across the resistor emphasizes the importance of placing a series impedance
between shunt non-linear devices. Without the series resistor, the spark gap
would probably never conduct.

If we continue this example, we find a current of about 1.8 A in the
zener diodes when there is 293 volts across the spark gap. This corresponds
to a worst-case power dissipation of about 8.5 watts in the reverse-biased
zener. There is reasonable coordination between the varistor and zener
diodes.

The 1N3825 zener diodes chosen by Smithson have a shunt capacitance of
about 2 nF. This is in addition to the 8.5 nF shunt capacitance of each
varistor. These large shunt capacitances will degrade the rise time of
digital signals and reduce the system bandwidth. The output of such a
transient protection circuit on a digital signal line should be connected to
an interface circuit with a Schmitt-trigger input to restore the proper logic
waveform.

Smithson (1977) placed the zener diodes on the same circuit board that
contained the load to be protected. This prevented ground impedance from
impairing the clamping voltage. The other protection components were located
at the point of entry for the cable into the building.

ANALOG APPLICATIONS

Integrated circuit operational amplifiers are an inexpensive, versatile analog "building block" which can be used to solve most analog signal conditioning problems for which the highest frequency of interest is less than about 1 MHz. Operational amplifiers are commonly found connected to analog inputs and analog outputs. Protection of these inputs and outputs condenses to the problem of protecting the operational amplifiers. We will consider how to protect inverting and non-inverting voltage amplifier inputs. Then we will discuss how to protect the output port.

INVERTING VOLTAGE AMPLIFIER

When the operational amplifier is used as an inverting voltage amplifier, electrical overstress protection for the input terminals is provided by the circuit shown in Fig. 11-6. The voltage gain of this circuit is $-R_f/R_{in}$. The values of R_f and R_{in} are usually between 1 kΩ and 1 MΩ, although R_f is sometimes larger than 1 MΩ.

The two diodes, D_1 and D_2, in Fig. 11-6 prevent the magnitude of the differential-mode input voltage from exceeding about 0.7 volts. Since the non-inverting input is directly connected to ground, these diodes also limit the common-mode input voltage. All commercially available operational amplifiers that do not have internal diode clamping can easily withstand ± 0.7 volts across the two input terminals. In fact, most modern integrated circuit operational amplifiers input terminals can be continuously connected to any potential between the power supply voltages (which are typically +15 volts and -15 volts) without damage to the operational amplifier.

OVERSTRESS

V_{in}

R_{in}

D_1, D_2

R_f

$-$

$+$

V_{out}
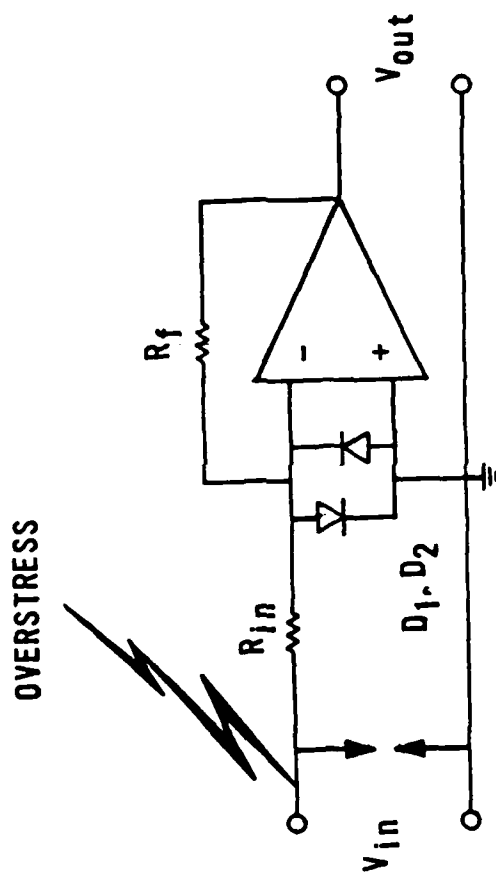
Fig.11-6

During normal operation of the operational amplifier there is no more than a few millivolts potential difference across D_1 and D_2, so the diodes will be non-conducting and effects of the diode's capacitance will be negligible. There is no hazard to D_1 and D_2 from operation in the reverse breakdown region since one of these diodes is always forward-biased and will protect the other diode from operation at a large reverse-bias potential difference.

Switching diodes (e.g. 1N4447, 1N4148 types) are commonly used for D_1 and D_2 because these diodes have fast response and are inexpensive. These diodes can tolerate steady-state currents of 500 mA without damage. This current corresponds to an input voltage of 500 volts if R_{in} is 1 kΩ, and an even larger input voltage if R_{in} is greater. If the spark gap shown in Fig. 11-6 has a DC firing voltage between 90 and 250 volts, the spark gap and resistor R_{in} will protect diodes D_1 and D_2 from excessive currents.

The weakest link in the circuit of Fig. 11-6 may be resistor R_{in}. It should be rated to withstand large transient voltages across it without degredation. (It is unacceptable to allow R_{in} to degrade since its value determines the voltage gain of the amplifier.) A 1 watt or 2 watt carbon composiition resistor is generally suitable, so is a 3 or 5 watt wirewound resistor. Metal film or carbon film resistors should be avoided for R_{in}. If continuous overstresses are anticipated, a PTC device should be inserted in series with an ordinary resistor for R_{in}. The PTC device will protect the ordinary resistor from damage from excessive power dissipation.

If the operational amplifier has a JFET, MOSFET, or other input stage

with input terminal currents of the order of a picoampere or less, then
standard switching diodes (e.g. 1N4447 type) are inappropriate for D_1 and D_2
if the very large input impedance of the amplifier is to be maintained. In
this situation one should use special low-leakage diodes for D_1 and D_2. Such
diodes can be fabricated from the gate to channel junction of a n-channel
JFET. If the larger forward conduction voltage is tolerable, one could use a
pair of GaAs green LEDs for D_1 and D_2. To maintain the low-leakage
conditions, the inverting input terminal should be connected to a PTFE
(Teflon, registered trademark of DuPont) insulated standoff. Low-leakage
diodes have maximum steady-state currents between 10 mA and about 100 mA. If
steady-state overload conditions are anticipated, the value of R_{in} should be
more than 1 kΩ and the spark gap should have a DC conduction voltage of 75 to
90 volts.

NON-INVERTING VOLTAGE AMPLIFIER

Sometimes one needs to protect a non-inverting voltage amplifier, Fig.
11-7a. Resistors R_1 and R_2 determine the voltage gain; typical values are:

$$10 \ \Omega \ \leq \ R_1 \ \leq 1 \ k\Omega$$

$$1 \ k\Omega \ \leq \ R_2 \ \leq 10 \ k\Omega$$

Since this circuit can not have a voltage gain less than one, the worst-case
maximum useful input voltage is equal to the maximum output voltage of the
operational amplifier. This is typically about 12 volts if ±15 volt power
supplies are used. If the voltage gain is greater than one or if the power
supplies are less than ±15 volts, the maximum useful input voltage is less
than 12 volts. For example, if the gain is 10, then the maximum useful input
voltage is only about 1.2 volts if ±15 volt power supplies are used and only
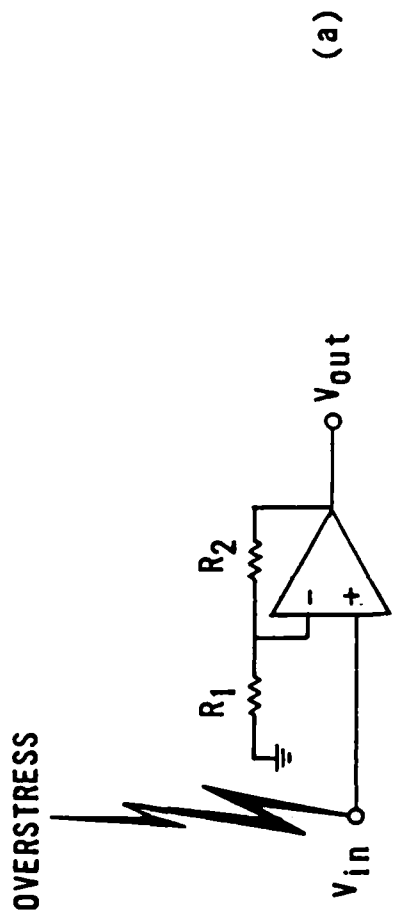
Fig.11-7
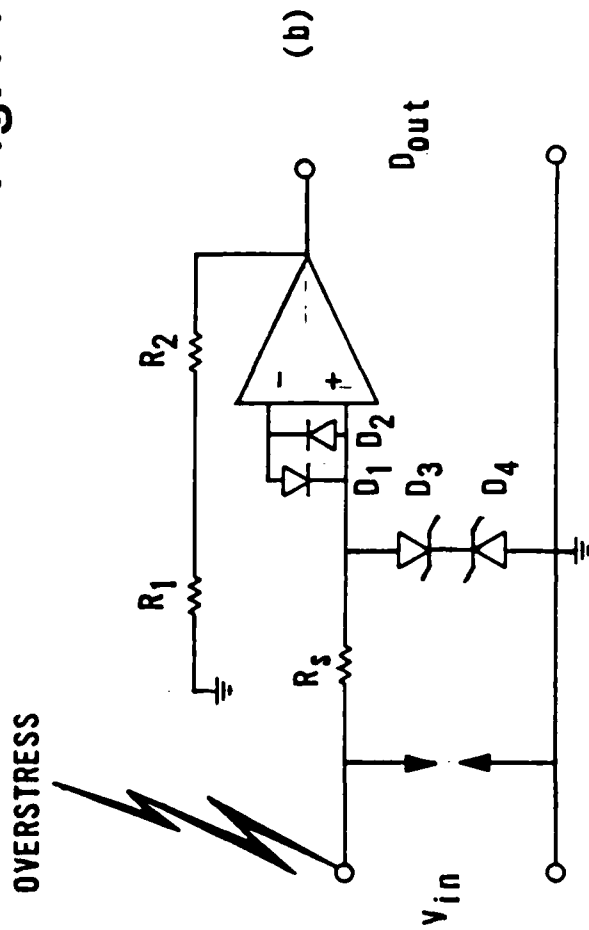
about 0.4 volts if ±5 volt power supplies are used.

A recommended protection circuit is shown in Fig. 11-7b. Diodes D_3 and D_4 are avalanche (or zener) diodes that are connected to form an absolute value clamp circuit. D_3 and D_4 conduct when the absolute value of the voltage across them exceeds V_z + 0.6 volts. This conduction voltage, V_z + 0.6 volts, should be slightly greater than the maximum useful input voltage so that D_3 and D_4 do not interfere with the normal operation of the circuit. If D_3 and D_4 were to conduct during normal operation, they would act as a voltage divider with resistor R_s. Moreover, operation of avalanche diodes at small currents (between about 1 µA and 1 mA) can cause the reverse-biased diode to become a broadband noise source.

However, values of V_z greater than 12 volts are not recommended. It is desirable to keep the voltage at the non-inverting input terminal of the operational amplifier _less_ than the positive power supply voltage. When ±15 volt supplies are used, this means the non-inverting input terminal should be kept at less than 15 volts from ground. There are several terms that can cause the clamping voltage to increase from the nominal value:

ΔV due to large current in reverse-biased diode	>2 volts
effect of 5% tolerance on 12 volt diode	0.6 volts
forward biased diode in D_3, D_4 pair	0.7 volts
parasitic inductance in the diode path	?

The values of ΔV due to large currents, and the forward biased voltage drop, were discussed in detail in Chapter 5. Because of these terms, one can not use diodes with Vz = 15 volts to clamp the input at 15 volts. There must be an interval of a few volts between V_z and the maximum clamping voltage.

When the maximum useful signal is between 1 and 6 volts, one may wish to specify $V_z = 6.8$ volts for D_3, D_4. Diodes with values of V_z less than about 6 volts use the Zener mechanism rather than the avalanche mechanism, and have much larger values of ΔV at large currents than avalanche diodes.

Diodes D_1 and D_2 prevent large differential-mode input voltages. Such voltages can arise if a rapidly changing signal or transient waveform is applied to the non-inverting input terminal so that the output voltage of the operational amplifier is slew-rate limited. These two diodes have no other function in this circuit.

The resistor R_s is included to limit the current to diodes D_3 and D_4 (and also D_1 and D_2 if R_1 is small). The input resistance presented by the operational amplifier circuit, r_{in}, is approximately

$$r_{in} = A_o \, r_i \, R_2/(R_1 + R_2)$$

where A_o is the open-loop voltage gain of the operational amplifier itself, r_i is the incremental resistance between the inverting and non-inverting input terminals. The value of A_o is typically at least 10^5, r_i is typically at least 1 MΩ, and $R_2/(R_1 + R_2)$ is usually between 1 and 10^3. This makes the input resistance of this circuit at least 10^8 Ω. Therefore, the value of R has negligible effect on the voltage gain for low-frequency signals when R is a few kilohms. However, significant attentuation of high-frequency signals can occur owing to the low-pass filter that is composed of R_s and the parasitic capacitance of avalanche (or zener) diodes D_3 and D_4. The time constant of this filter is usually of the order of a few microseconds since R_s

~ 1 kΩ and C ~ 3 nF. This may not be a serious limitation since common, inexpensive operational amplifiers (e.g. 741 and 307 types) also act as a low-pass filter with a time constant of 1 μs due to the gain-bandwidth product of the amplifier.

To reduce the value of the RC time constant formed by the protective circuit, one could use a PTC device in series with (or in place of) an ordinary resistance. During normal operation, the value of R could be as small as 22 Ω. The capacitance of the clamping circuit (D_3, D_4 in Fig. 11-7b) could be reduced by using a series string of switching diodes in place of the avalanche diodes, as described in Chapter 6, or in series with the avalanche diodes, as described in Chapter 5.

PROTECTION OF OUTPUT

Destructive overvoltages can also reach the operational amplifier through the output port. The circuit shown in Fig. 11-8 is suggested for protection of the operational amplifier from overstresses on an output cable. Diodes D_5 and D_6 can be avalanche diodes with V_z ~ 12 volts when the operational amplifier is operated from ± 15 volt supplies. When smaller supply voltages are used, the value of V_z should be chosen to be near the maximum output voltage of the operational amplifier. Notice that no resistance is included between the operational amplifier output terminal and D_5, D_6. If these diodes conduct when the magnitude of the output voltage is large, we rely on internal output current limiting in the operational amplifier (which is also called "short-circuit protection") to protect the operational amplifier, as well as diodes D_5 and D_6. If the maximum output current from the operational
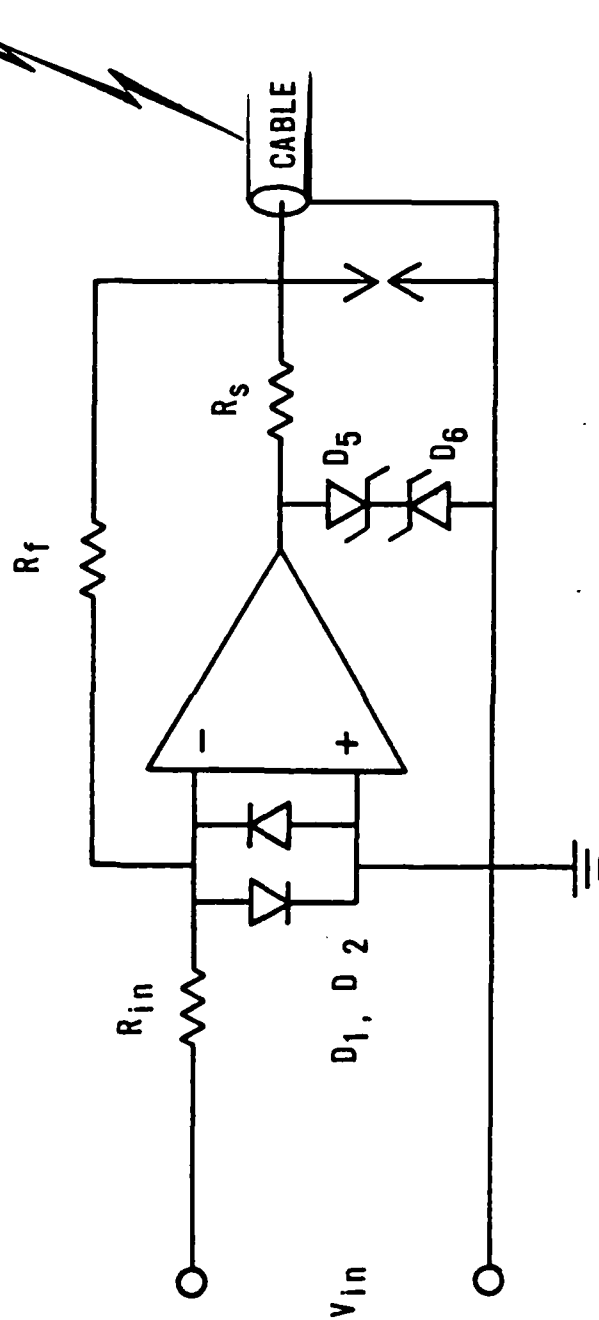
Fig. 11-8

amplifier could harm D_5, D_6, or if there is no output current limit inside the operational amplifier, then a PTC device should be included between the operational amplifier output and D_5.

The resistor R_s limits the current in D_5 and D_6 from overstresses that appear on the output cable. Because the output current from the operational amplifier must also pass through R_s, we can not automatically specify a large resistance (e.g. 1 kΩ) at this point in the circuit. The maximum output current of a typical integrated circuit operational amplifier is about 15 mA. If the maximum acceptable voltage drop across R_s during normal operation is 1 volt, R_s must be less than 68 Ω. A PTC device is desirable, instead of an ordinary resistance, for R_s. If an ordinary resistance is used for R_s, a value of 330 Ω is suggested in order to protect the avalanche diodes. If large transients are anticipated and R_s is an ordinary resistance, then a 2 watt carbon composition or a 3 to 5 watt wirewound unit is recommended.

Notice in Fig. 11-8 that the feedback point is not taken at the output terminal of the operational amplifier. The circuit shown in Fig. 11-8 has a small output impedance, when viewed from the cable, despite the large value of R_s. This surprising result is produced by the negative feedback and the choice of the feedback point shown in Fig. 11-8.

Because it is possible for overvoltage on the output cable to reach the inverting input terminal through resistor R_f, the input must also be protected. Comprehensive input protection is furnished by switching diodes D_1 and D_2 when R_f is at least 1 kΩ. Typical values of R_f in this circuit would be of the order of 10 kΩ, since smaller values of R_f divert output current

away from the output cable to the input circuit.

LOGIC APPLICATIONS

Protection of input and output terminals of integrated circuit logic devices is particularly challenging. Faster logic devices are more vulnerable to damage from electrical overstresses, for reasons specified in Chapter 1. Yet circuit designers, for good reasons, have gone away from the slow, robust DTL and ordinary TTL logic families, and tended to use various versions of low-power Schottky TTL which are much more vulnerable to damage.

Logic circuits commonly use a single +5 volt power supply, although some CMOS circuits can operate with power supply voltages as small as +3 volts. It is difficult to obtain protection devices that will clamp a line at 3 to 5 volts during a transient. The best device in common usage is an avalanche diode with a value of V_z = 6.8 volts. True zener diodes have a relatively large value of $\Delta V/\Delta I$ and do not offer tight clamping.

OPTOISOLATOR FOR PROTECTION OF DIGITAL DATA LINES

The circuits described above can be used to protect either line drivers or line receivers, since there is nothing in the protection circuit to restrict the flow of information to a single direction. This situation changes when optoisolators are used. The information flow must pass from the LED side to the photodetector side. This introduces the complication that separate circuits must be designed for receivers and transmitters. In addition, the photodetector output usually requires an active circuit to

condition the signal. The protection circuit for the transmitter is particularly complicated, as we will see. In the following examples, circuits that contains an optoisolator will be described for receiving and sending computer data according to the RS-232 interface specifications.

The RS-232 data communications standard was never intended for use on cables with a length greater than about 15 metres (50 feet). The RS-232 system is "single-ended," that is, there is one common conductor for all of the bidirectional signal conductors. This presents a problem when the two ends of the cable are in different buildings, where there may be a difference of potential between "ground" at the two buildings. The major symptom of this ground loop is a quasi-sinusoidal 60 Hz waveform superimposed on the digital data transmission. Such a problem has been avoided in the newer RS-423 data communication standard by requiring receivers with differential inputs and having a separate common conductor for signals that propagate each way. The RS-422 data communication standard takes an even more aggressive approach by using a balanced pair of wires for every digital signal, so that there is no common conductor, and also requires receivers with differential inputs. However, advantages of these newer standards are not readily available to users of computer terminals and peripherals, since continued use of RS-232 makes it easy to interface older equipment to newer equipment. By incorporating isolation into the active protection circuit, we can interrupt the ground loop that is inherent in RS-232 communication circuits.

ACTIVE CIRCUIT FOR RECEIVED DATA

Fig. 11-9 shows a schematic of an improved transient isolation circuit

ACTIVE RECEIVER
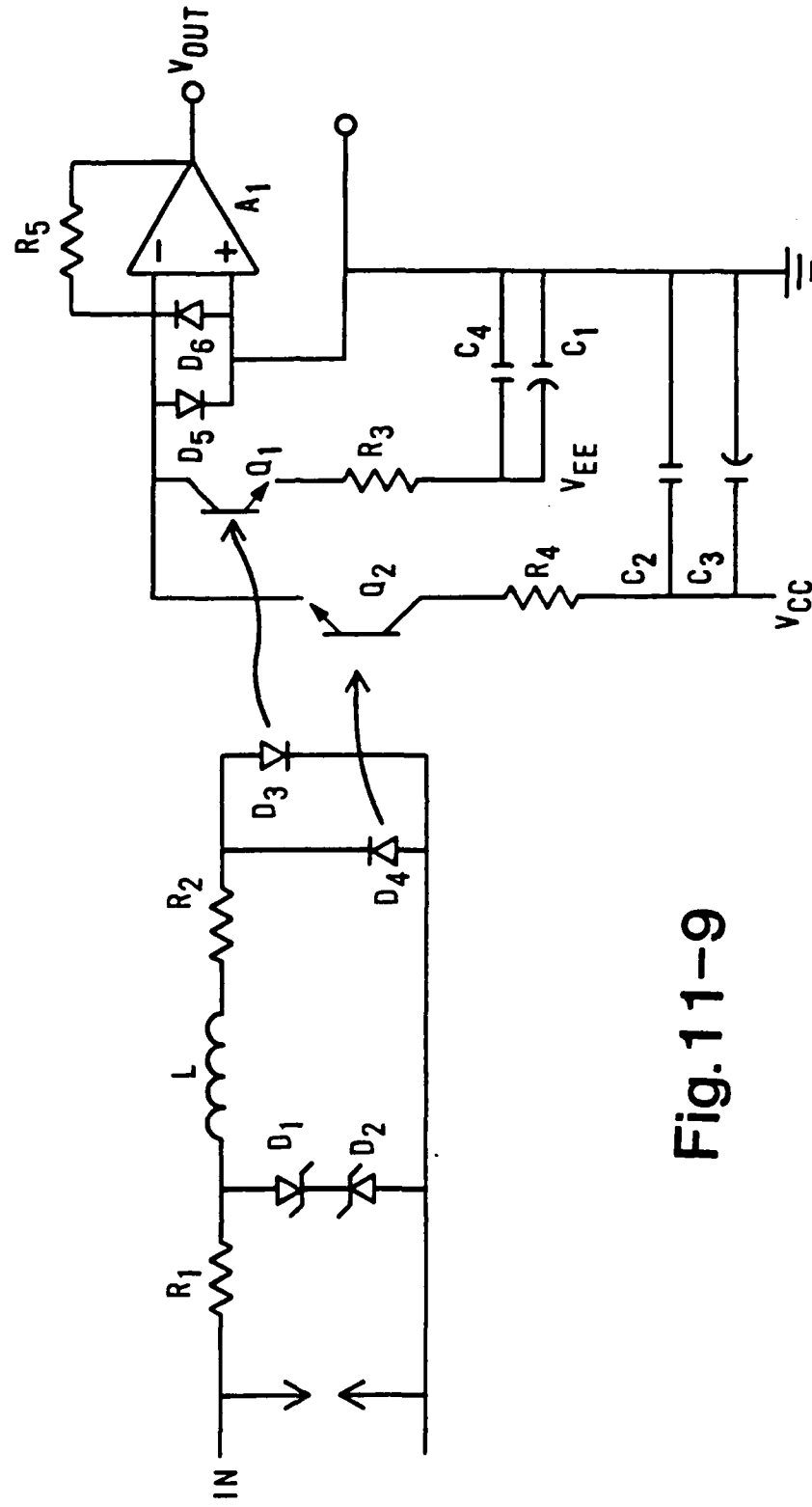
Fig. 11-9

that, in addition to the features of the circuit shown in Fig. 11-1, also provides protection against common-mode transients. Values of components are given in Table 11-I. The common-mode protection is provided by two optical isolators (D_3 and Q_1; D_4 and Q_2) that combine light emitting diodes (LEDs) and phototransistors in the same package. This optoisolator should be rated by the manufacturer to withstand at least 3000 volts between the LED and the phototransistor, be relatively fast, and have a large phototransistor gain.

The part of the circuit in Fig. 11-9 that is composed of a spark gap, resistors R_1 and R_2, and avalanche diodes D_1 and D_2, is similar to that in Fig. 11-1. The series combination of resistors R_1 and R_2 provides 3000 ohms impedance, the minimum allowed by RS-232C and limits the current in the LEDs. The maximum current in the LEDs will be no more than about 4.7 mA since the infrared LED has a drop of about 1 volt and the maximum input signal 15 volts. The minimum non-zero LED current, 1.3 mA, occurs when the voltage on the RS232 line is 5 volts, the minimum specified magnitude. The minimum current generates sufficient light in the LED to cause the phototransistor to conduct. The light emitting diodes are connected back to back (anti-parallel) so that one LED is always forward-biased. The forward voltage drop (about 1 volt) of one LED prevents the other LED from being operated in the reverse breakdown region (more than 3 volts). Because the RS232 signal is bipolar, two optoisolators are required: one for each polarity.

The maximum magnitude of an RS232 signal is 15 volts when the output is terminated with a resistance between 3000 and 7000 ohms. The largest signal voltage that will appear across avalanche diodes D_1 and D_2 is 5.7 volts (notice that R_1 and R_2 form a voltage divider and recall that the LED has a 1

volt forward drop). The value of R_1 in Fig. 11-9 can be larger than the value

of R in Fig. 11-1 because the Fig. 11-9 circuit terminates a RS232 line.

Therefore there is no particular need for avalanche diodes D_1 and D_2 in Fig.

11-9 to have large power ratings; 1 watt models are sufficient. One

manufacturer rates its one watt steady-state 1N4736 avalanche diodes at 60

watts for a 15 µs non-repetitive rectangular pulse (Motorola Zener Diode

Manual, 1980, page 11-56). This is adequate overload capability, since 2 watt

dissipation in the avalanche diode will produce more than 550 volts across the

spark gap (which is assumed to be non-conducting initially) when R_1 = 2000

ohms. The spark gap will quickly conduct and the power dissipated in the

avalanche diodes will then be much less than their maximum steady-state value.

During a transient, avalanche diodes D_1 and D_2 will limit the magnitude of the

voltage to about 10 volts. Resistor R_2 will then limit the current in the

forward biased LED to about 10 mA, well below the 90 mA which is the maximum

continuous current specified by the manufacturer.

The output of the optoisolators are recombined with an operational

amplifier. When either of the phototransistors is on, the operational

amplifier is in a non-linear mode and the inverting input is not a virtual

ground. Larger values of R_3 and R_4 than those given in Table 11-I increase

the switching times of phototransistors Q_1 and Q_2. The operational amplifier,

A_1, should be moderately fast (between 6 and 30 V/µs slew rate to satisfy

sections 1.3, 2.3, and 2.7 of RS232). Diodes D_5 and D_6 protect the operational

amplifier input stage from large input voltages and allow the phototransistor

collector current to flow to ground. If the optoisolator should fail, these

diodes should protect the operational amplifier (and the load).

The operational amplifier has internal protection against output short-circuits to any voltage between the +15, -15 volt supplies, including ground. The operational amplifier alone provides a power-off source impedance of much more than the 300 ohms that is specified in section 2.5 of RS-232C.

Bypass capacitors C_1, C_2, C_3, and C_4 are included to reduce the effect of inductance in the power supply lines.

ACTIVE CIRCUIT FOR TRANSMITTED DATA

A circuit that provides isolation and overvoltage protection for a line driver that transmits data over a long cable is shown in Fig. 11-10. Values of the components shown in Fig. 11-10 are listed in Table 11-II.

Protection against common-mode transients is provided by two optical isolators (D_9 and Q_3; D_{10} and Q_5). Transistors Q_4 and Q_6 are NPN power transistors that are more robust than output transistors in integrated circuits or in optoisolators.

Resistor R_{18} and optoisolator LEDs, D_9 and D_{10}, form the entire receiving circuit which is completely isolated from local ground and the long line. The value of R_{18} is 3000 ohms, the minimum allowed for a RS232 compatible receiver. This choice of R_{18} maximizes the LED current, which is desirable because we want to drive either Q_3 or Q_5 into saturation.

NORMAL OPERATION OF TRANSMITTER

Fig. 11-10 ACTIVE TRANSMITTER
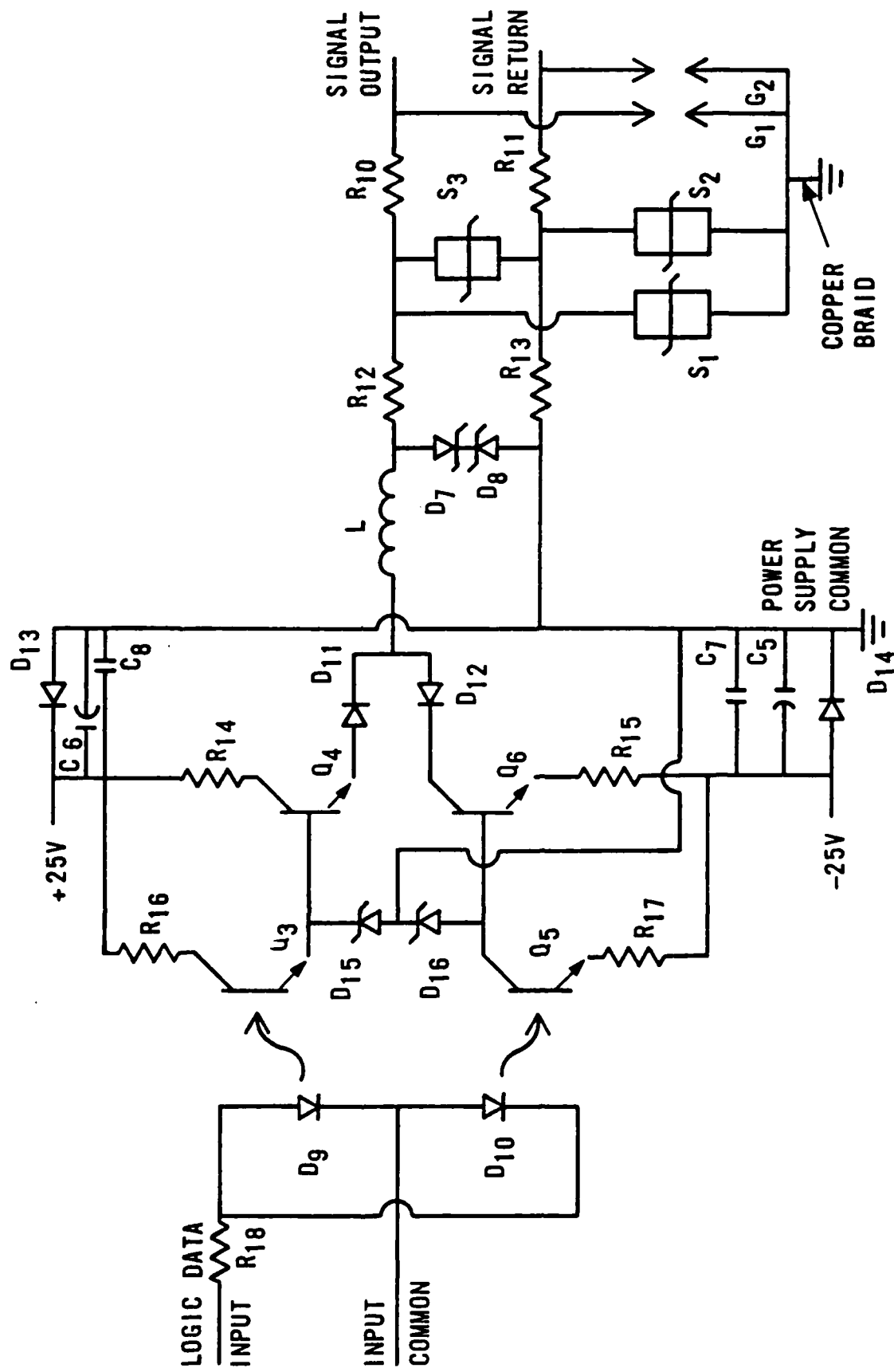
Before we analyze the behavior of the circuit under stress, let us consider the normal operation. Under normal conditions the transient protection devices are non-conducting and the circuit shown in Fig. 11-10 can be simplified to Fig. 11-11.

Suppose the data line is positive so D_9, Q_3, and $Q4$ are all on and D_{10}, Q_5, and Q_6 are all off. We shall consider only the positive half of the circuit (upper half of Fig. 11-11) since the circuit is symmetrical.

The optoisolator characteristics and value of R_{16} should be chosen so that Q_3 supplies at least 1 mA current to D15 and the base of Q_4 when the input voltage across R_{18} and D_9 is only 5 volts, the minimum RS232 level.

The avalanche diode D_{15} establishes a constant voltage between the base of Q_4 and the output common wire. Transistor Q_4 is operated in the active region as a series-pass voltage regulator. The output voltage is about 14.8 volts and the collector-emitter voltage of Q_4 is about 9 volts.

If the termination resistor, R_T is 3000 ohms, then the steady-state collector current of Q_4 is about 4.8 mA. The series resistor R_{14} serves to limit the maximum collector current in Q4. We can make R_{14} - 68 Ω, without appreciably increasing the output impedance from the case where R_{14} is zero. The maximum collector current, if the output signal line were shorted to the power supply common and Q_4 were operated in saturation, would be about 140 mA. This current will not harm Q_4.

Resistances R_{11} and R_{13} may be shorted by the resistance of the earth

Fig.11-11
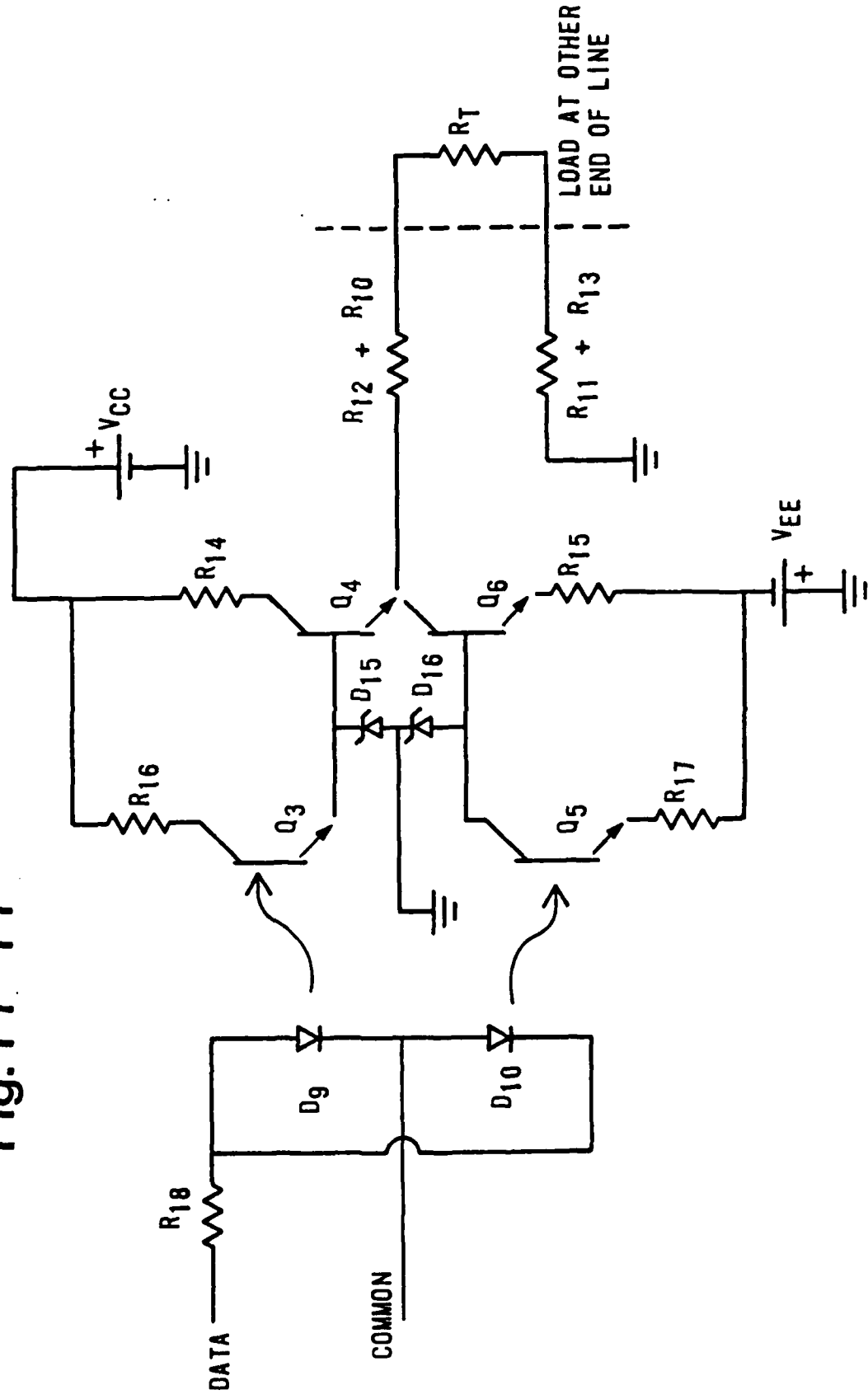
between the transmitter site and the receiving site (where R_T is located). During normal operation ground potential at both sites should be nearly identical so most of the signal return current may not pass through R_{11} and R_{13}.

The values for resistors R_{10} through R_{13} are determined by transmission line considerations. It would be desirable for the output impedance of the transmitter to match the characteristic impedance of the cable. The twisted-pair cable probably has a characteristic impedance between about 100 Ω and 150 Ω, thus we make $R_{10} + R_{12} = 100$ Ω. R_{10} is a positive temperature coefficient resistor, which protects the avalanche diodes and varistors from sustained currents.

SPARK GAP AND VARISTORS

We now consider the transient protection devices in Fig. 11-10. The spark gaps G_1 and G_2 keep the potential of the long lines within about 150 volts of local ground. These two spark gaps can be contained in a single three-electrode gap, or they can be independent two-electrode gaps. Since this is not a balanced line driver, the three-electrode gap is not essential and we can obtain a smaller firing voltage by specifying two-electrode gaps. Notice that a spark gap, G_2, is used between local ground and the signal return conductor in the long line. When lightning strikes a building there can be large potential differences between ground in that building and adjacent buildings. These potential differences are caused by the finite conductivity of the earth. There is no a priori reason to assume that the transient voltage will be larger (or smaller) on the output signal conductor

as compared to the signal return conductor. Therefore, both conductors are shunted to local ground with a spark gap.

The finite time required for a breakdown to occur inside the spark gap increases the stress on the remainder of the circuit. Therefore metal oxide varistors have been included between the spark gaps and avalanche diodes. One must include resistors, R_{10} and R_{11}, between the varistors and spark gaps if the spark gaps are to protect the varistors from large current surges. When a transient current of 10 A passes through the varistor, there will be about 270 volts across the spark gap with the components listed in Table 11-II. The spark gap will conduct within a few microseconds (or less), and shunt current away from the varistor.

Whenever the spark gap G_1 conducts and operates in the arc region, the voltage across the spark gap will be less than the 47 volt value of V_N, and varistor S_1 will not conduct. The same situation is true for the spark gap G_2 and the varistor S_2.

Varistor S_3 limits the voltage difference between the lines. Without varistor S_3, a larger current could flow in resistors R_{12} and R_{13}, and in diodes D_7 and D_8.

Avalanche diodes D_7 and D_8 limit the maximum magnitude of collector to emitter voltage across power transistors Q_4 and Q_6. The minimum value of avalanche voltage would be 15 volts, the maximum signal output voltage. The maximum avalanche voltage would be determined from the maximum allowable collector-emitter voltage of Q_4 and Q_3 (and Q_5 and Q_6). The breakdown

voltage, V_{CEO}, is at least 70 volts for Q_3 and Q_5 if they are part of a CNY17 optoisolator. The V_{CEO} breakdown value of Q_4 and Q_6 is even greater. Thus avalanche diodes D_7 and D_8 must conduct at less than 45 volts to keep Q_3 and Q_5 from operating in the breakdown region. The maximum allowable current for a given type of avalanche diode is greater for smaller values of avalanche breakdown voltage. The choice of a small avalanche voltage also reduces the stress on the transistors Q_3, Q_4, Q_5, and Q_6.

When the emitter terminal of a NPN transistor is about 7 volts more positive than the base terminal, the base-emitter junction is in the reverse breakdown region. Diodes D_{11} and D_{12} protect the transistors from operation in this region.

TRANSIENT ANALYSIS

First we consider the largest pure differential mode voltage that will not cause the varistors or spark gaps to conduct. For the devices listed in Table 11-II, this situation occurs when the output signal conductor is about +27 volts from local ground and the signal return conductor is about -27 volts from local ground. An current of about 90 mA flows through R_{10}, R_{12}, D_7, and D_8. An initial current of about -260 mA flows through R_{11} and R_{13}. (We say "initial" because the PTC resistor, R_{10} will switch and decrease this current.) A current of about 170 mA flows in the power supply common wire when the transient voltage is purely differential-mode and avalanche diodes D_7 and D_8 are conducting.

We have just considered the case for which the transient is purely

differential-mode. Now we consider a worst-case purely common-mode transient for which the varistors or spark gaps are non-conducting. Suppose both the signal output and the signal return conductor are 55 volts above local ground. The initial current in R_{10}, R_{12}, D_7, and D_8 is 360 mA; the initial current in R_{11} and R_{13} is 530 mA. The current in the power supply common wire is the sum of these two currents, which is about 0.9 A. It is desirable to avoid such a large current in the common wire since this current will produce voltage drops due to resistance and inductance in the wire.

We can reduce the effect of these large currents by specifying that the common wire be relatively large (e.g. 16 AWG or larger). Notice in Fig. 11-10 that the spark gaps and varistors have a separate ground path. Copper braid with a width of at least 5 millimetres would be preferable to wire. The spark gap ground braid and power supply common wire are connected to the same earth ground (i.e. copper water pipe or building ground), but are routed independently so that the spark gap and varistor currents have a minimal effect on the power supply voltages.

Smaller varistor voltages would reduce the size of these common currents, but that choice would also make it less likely that the spark gaps would conduct. We could also reduce the size of the common currents by increasing the value of R_{10} through R_{13}, but this would also increase the output impedance of the circuit.

After the positive temperature coefficient resistors, R_{10} and R_{11}, switch, the current in the varistors, avalanche diodes, and power supply common wire will be greatly reduced. This protects these components from

sustained overstresses.

POWER SUPPLIES

The subject of DC power supplies is treated in detail in Chapter 13. We make only a few remarks here that are relevant to this particular application.

The power supplies were chosen to be $\pm$ 25 volts, the maximum allowable open-circuit RS232 voltage. Once a transient has entered the power supply it could damage many amplifiers and the equipment that is connected to these amplifiers. Therefore these 25 volt supplies for the transmitter circuit shown in Fig. 11-10 should be independent of the $\pm$15 volt supplies used in the receiver, which was shown in Fig. 11-9. However, the same 25 volt supplies can be used for many different active transmitter circuits. The active transmitter is robust and can tolerate power supply transients.

The $\pm$ 25 volt power supplies are protected and isolated from other transmitters by resistors R_{14} through R_{17}, and capacitors C_5, C_6, C_7, and C_8. The diodes D_{13} and D_{14} protect the polar capacitors from possible voltage reversals during transients. Additional protection for the power supplies should be provided by a varistor across the output terminals of the power supply chassis.

The insulation of the step-down transformer that furnishes power to the $\pm$25 volt regulators should be rated to withstand at least 2500 volts between primary and secondary. The capacitance between primary and secondary coils of this transformer should be less than about 50 pF. Otherwise the system

isolation provided by the optoisolator might be compromised by breakdown of the transformer insulation (and injection of transient current into the power line). Metal oxide varistors should be connected at the transformer primary winding, as described in Chapter 12. These measures serve to protect the circuit from transient overvoltages on the power line and also serve to prevent the transient from propagating from the signal conductor to the mains.
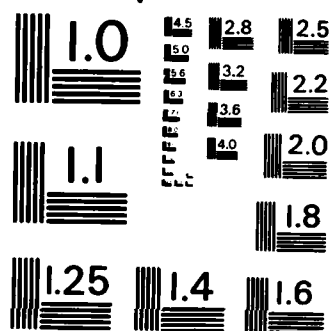
MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

Table 11-I

Components for Optoisolator Received Data Circuit

G_1 150 V DC spark gap in ceramic case

D_1,D_2 1N4736 (6.8 V, 1 watt Zener diode)

D_3,D_4 part of CNY17-IV optoisolator

D_5,D_6 1N4447 switching diode

Q_1,Q_2 part of CNY17-IV optoisolator
                     breakdown voltage $V_{CEO} > 70$ volts,
                     about 5 µs switching time between active and cutoff regions,
                     current transfer ratio > 160 %

A_1 LF356 operational amplifier

R_1 2000 ohms, 2W carbon composition

R_2 1000 ohms, 0.5W carbon composition

R_3,R_4 2200 ohms, 0.5W carbon composition

R_5 100 kilohms, 0.25W

C_1,C_2 100 µF, 25V

C_3,C_4 0.1 µF ceramic

Table 11-II

Components for Optically-isolated Transmitted Data Circuit

G_1, G_2 — 90 V DC spark gap in ceramic case

S_1, S_2, S_3 — metal oxide varistor, $V_N = 47$ volts

D_7, D_8 — 1.5KE18C (18V bipolar avalanche diode)

D_9, D_{10} — part of CNY17-IV optoisolator

$D_{11}, D_{12}, D_{13}, D_{14}$ — 1N4004 rectifiers

D_{15}, D_{16} — 1N5246 (16V, 0.5W avalanche diode)

Q_3, Q_5 — part of CNY17-IV optoisolator

Q_4, Q_6 — 2N6552   1A, 80V power transistor

R_{10}, R_{11} — positive temperature coefficient resistor; each about 20 Ω during normal operation, switch at about 130 mA in 20 celsius environment.

R_{12}, R_{13} — 82 ohms,   2.0 W   carbon composition or wirewound

R_{14}, R_{15} — 68 ohms,   2.0 W   carbon composition

R_{16}, R_{17} — 2200 ohms, 0.5 W   carbon composition

R_{18} — 3000 ohms, 0.5 W   carbon composition

C_5, C_6 — 1.0 µF to 4.7 µF, tantalum, 35 V

C_7, C_8 — 0.1 µF, ceramic, 50 V

PROTECTION OF MAINS

INTRODUCTION

Transient suppression devices for mains (AC power lines) are appropriately divided into four classes.

1. Distribution surge arresters (typically 3 kV_{rms} and greater)

2. Secondary arresters (120, 240, 480 V_{rms})

3. Line conditioners, surge protection modules (120 V_{rms})

4. Surge protection devices inside a chassis

These classes can be distinguished in several ways: type of circuit and construction, electrical surge specifications, and who is responsible for requesting it.

The local power company installs and maintains the distribution arresters. These high-voltage devices will not be discussed further in this report. Additional information on such devices can be found in the textbook by Alan Greenwood (1971) and in proprietary literature from manufacturers (e.g. General Electric, Westinghouse, McGraw-Edison).

The secondary arrester is installed by an electrician at the point of entry of power into a building. If the building contains a distribution transformer, the secondary arrester (as the name implies) is connected on the secondary side (120 volt rms side) of the distribution transformer. If the building does not contain a distribution transformer, the secondary arrester is usually mounted on the meter box or on a disconnect box at the service entrance. A circuit diagram is shown in Fig. 12-1. The occupants of the

DISTRIBUTION POWER LINE

BLACK

WHITE

BLACK

EARTH GROUND AT TRANSFORMER

BLACK

GREEN

WHITE

BUILDING GROUND AT POINT OF ENTRY

BLACK

SECONDARY ARRESTER AT POINT OF ENTRY
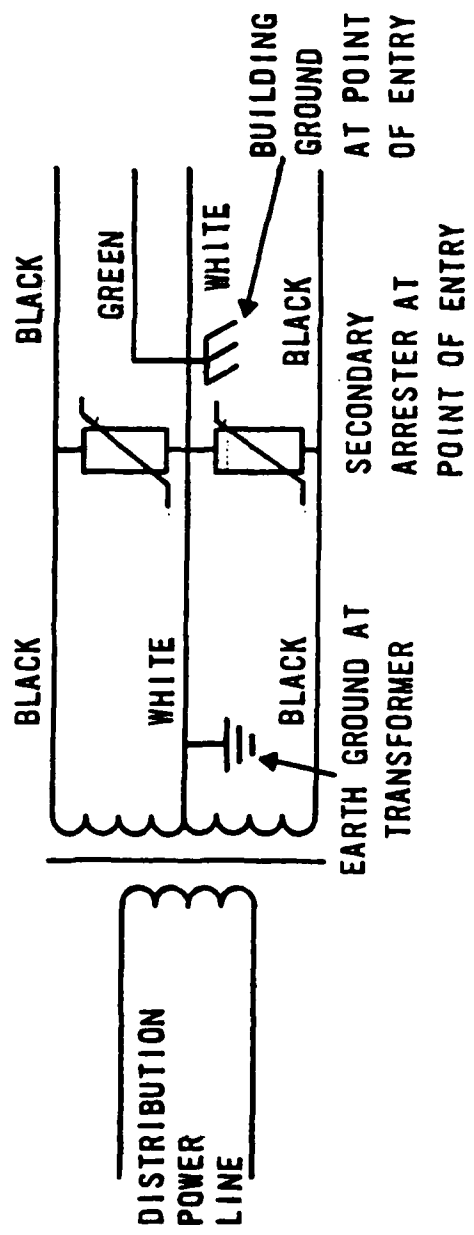
Fig.12-1

building are responsible for requesting the installation of a secondary arrester.

The third class of devices, surge protection modules and line conditioners, are boxes that are connected to a wall outlet to furnish "transient-free" power. A circuit diagram is shown in Fig. 12-2. The user simply plugs the load that is to be protected (e.g. computer) into the output of the box. Whoever requests the sensitive electronic circuit is responsible for also obtaining "transient-free" power, most commonly, by requesting a surge protection module or line conditioner. Line conditioners are a device that is designed to maintain the rms value of line voltage at a nearly constant value, as well as provide isolation and remove noise. The surge protection module usually contains a metal oxide varistor(s), and perhaps also an RFI filter circuit. We shall discuss line conditioners and surge protection modules in greater detail later.

The fourth, and last, class of AC power transient protection devices are those that are included inside the chassis of electronic equipment by the manufacturer. Such devices include (1) passive RFI/EMI filters in pre-fabricated metal cases and (2) a metal oxide varistor that is connected across the mains. We shall discuss appropriate specifications for these devices later.

ISOLATION TRANSFORMER

BRANCH ARRESTER

15A CIRCUIT BREAKERS

60A MAIN BREAKER

BLACK

WHITE

GREEN
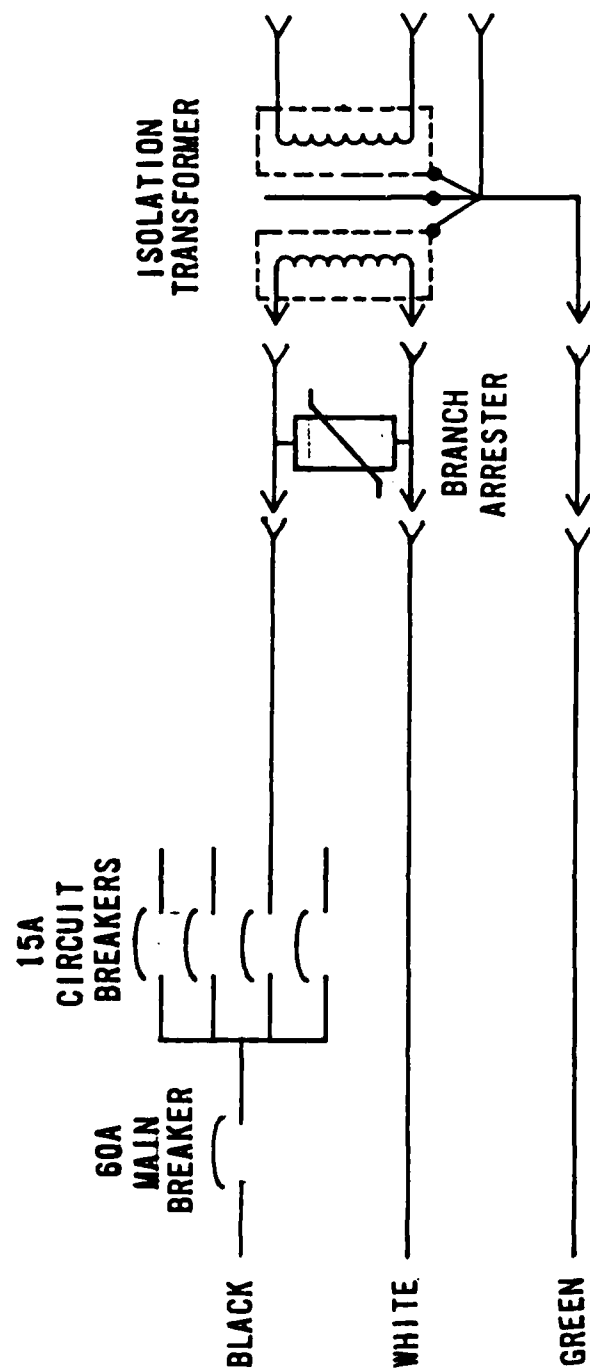
Fig. 12-2

SECONDARY ARRESTER

Secondary arresters are typically rated for 500 to 650 V_{rms} service so that one model is suitable for use on 120, 240, or 480 V_{rms} lines. There are two common kinds of secondary arresters. The older kind has a silicon carbide varistor in series with a spark gap; the newer kind contains a metal oxide varistor alone. One side of the device is connected to the line, the other side is connected to local ground. One, two, or three devices are available in a single container for protecting single phase, two phase, or three phase secondary circuits. A circuit diagram was shown in Fig. 12-1.

Military Standard 188-124 (1978, section 5.1.1.3.12) for long haul/tactical communication systems requires a lightning arrester to be installed at the point of entrance of the mains into the facility.

The silicon carbide/spark gap units are designed to clamp the power line at about 2.5 kV @ 1.5 kA and at about 4 kV @ 10 kA. The spark gap is set to conduct at about 2 kV. The silicon carbide varistor is included to extinguish the spark gap and has an approximate characteristic curve

$$I = 4.5 \times 10^{-11} V^4$$

where I is in amperes and V is in volts.

The metal oxide varistor units clamp at a somewhat smaller voltage, about 2.9 kV @ 10 kA. A typical value of V_N for a 120 volt rms secondary arrester is about 1 kV. Such a large value of V_N is necessary to prevent excessive

steady state power dissipation (and consequent destruction of the varistor) after degredation by transient overstresses.

Applications in which the series combination of a <u>metal</u> <u>oxide</u> varistor and a spark gap are connected across the mains seem pointless. We can not use the spark gap alone across the mains, owing to follow current problems, which the varistor solves. However, why not use the metal oxide varistor alone? The spark gap in no way helps the varistor to survive large transient currents. Moreover, the finite turn-on time of the spark gap may allow fast transients to propagate downstream from the arrester that contains a spark gap. We note that the series combination of a <u>silicon</u> <u>carbide</u> varistor and a spark gap <u>is</u> <u>useful</u> for mains protection. The spark gap acts as a switch to prevent steady-state power dissipation in the silicon carbide varistor during normal conditions and the varistor extinguishes follow current in the spark gap after a transient. In general, the metal oxide varistor will provide superior protection compared to the silicon carbide varistor/spark gap combination.

COORDINATION OF TPDs FOR MAINS

Because the clamping voltage of secondary arresters is so great, one also needs additional transient protection downstream (between the secondary arrester and the sensitive load). The reason for using the secondary arrester is that it will tolerate the large peak currents (typically 20 kA peak) that are encountered during a direct lightning strike to an overhead power line. It is not possible with present technology to fabricate an <u>economical</u> and reliable arrester that will tolerate 20 kA peak currents <u>and</u> clamp at a

relatively small voltage of, for example, 400 volts. Therefore the secondary arrester is designed to clamp at several kilovolts in order to keep the arrester inexpensive and maintenance-free. Since a 3 kV transient on the mains can damage electronic systems, additional protection devices are required downstream. Branch circuit TPDs limit the peak voltage on the mains to between about 300 and 400 volts. TPDs inside the chassis provide redundant protection from external transients and prevent a particular system from polluting the branch circuit with transients that originate inside the chassis.

The mains wiring should be enclosed in rigid steel conduit for optimal shielding against introduction of transients in the interior of the building (Lasitter and Clark, 1970, p.4).

AC LINE CONDITIONERS

An AC line conditioner is a box that contains a circuit to accomplish all of the following three items:

1. Provide voltage regulation. When the rms input voltage is between 95 and 130 volts, the output voltage should be 120 ± 5 volts rms for no load to full load.

2. Provide at least 50 dB differential-mode attenutation at frequencies above 100 kHz

3. Provide isolation: no more than 1 pF capacitance between the input

and output terminals. This, when combined with capacitance shunted
across the load, attenuates common-mode transients.

We will now discuss how the voltage regulation is obtained, and then
discuss differential-mode attenuation. Common-mode attenuation was discussed
in Chapter 9 under isolation transformers.

There are two common techniques to provide AC voltage regulation in line
conditioners: (1) the ferro-resonant transformer circuit and (2) the tap-
switching transformer.

The physics of operation of a ferro-resonant transformer circuit has been
reviewed by Grossner (1983, p.160-162). The following description is grossly
oversimplified, but serves to convey the general idea. The ferro-resonant
transformer uses a core that is operated in "saturation," i.e. the magnitude
of magnetic induction, B, is essentially independent of the magnitude of the
magnetic field, H. Since the magnitudes of the input voltage, current in the
primary coil, and the magnetic field, H, are all proportional, this makes the
value of B essentially independent of the rms input voltage. If a resonant
circuit were not present, the output voltage would be a crude square wave with
the same frequency as the input voltage. An inductor-capacitor resonant
circuit converts the output voltage to a quasi-sinusoidal waveform. If a
sinusoidal output waveform is desired, a "harmonic-neutralized" design should
be specified. This will typically provide less than 3 % harmonic distortion.

The ferro-resonant circuit has several outstanding advantages compared to
other types of regulators:

1. Excellent voltage regulation: output voltage is typically 120 ± 3 volts for input voltages between 95 and 138 volts (no load to full rated load conditions).

2. The ferro-resonant transformer is inherently short-circuit proof. If the output terminals are shorted, the magnitude of the output current will be about 1.5 to 2.0 times the maximum rated load current. Under these conditions the transformer will act as a constant current source. This will not harm the transformer.

3. The ferro-resonant circuit tends to ignore brief losses of input power. There is essentially no change in rms output voltage when the input voltage is zero for durations of 2 to 4 ms, since the resonant circuit continues to oscillate for several cycles without additional energy input.

4. High reliability. The only components in a ferro-resonant transformer are one transformer (with multiple windings) and one capacitor. There are no moving parts, and no semiconductors.

However, the ferro-resonant transformer has several major disadvantages:

1. Because a resonant circuit with a fixed resonance frequency is used, the device is sensitive to changes in frequency of the input waveform. For a typical ferro-resonant transformer, if the input frequency deviates from the design frequency, the output voltage

will change by about 1.5% to 2% for each 1% change in input
frequency. This effect is inherent in the performance of a resonant
circuit that is not driven at the resonance frequency. This is not
a serious problem when the input power is obtained from public
utilities. However, when the input power is obtained from local
generators that are driven by an internal combustion engine, the
error in input frequency is often several percent.

2. Because the transformer core is driven into saturation, the
    transformer can be inefficient. If the ferro-resonant transformer
    is operated at half its maximum rated load, the typical efficiency
    is only about 65% due to large losses in the core. (If the unit is
    operated with its rated load and with an input voltage that is
    approximately the same as the output voltage, typical efficiencies
    are between 80 and 95%. These figures are often cited by vendors as
    evidence of "good efficiency.") The ferro-resonant transformer core
    operates at 45 to 85 celsius above ambient temperture due to power
    dissipated in the core. This heat burden can be a serious
    consideration in computer rooms that need cooling.

3. The ferro-resonant transformer is massive. A 500 VA ferro-resonant
    unit has a mass of about 21 kg, about 75% more than for a tap-
    switching line conditioner of the same VA rating.

A second type of AC voltage regulator is the tap-switching transformer.
A standard two coil (primary and secondary) transformer can be provided with
multiple terminals (called "taps") on either coil to compensate for variations

in the magnitude of the input (primary) voltage. Depending on the number of taps, one can obtain arbitarily good regulation. A common specification is for the output voltage to remain constant within ±5% for an input voltage change of ±12%. While this is worse regulation than that of a ferro-resonant circuit, it is still acceptable for most critical applications (e.g. computers, electronic instruments). The tap switching is usually done with a triac, which is faster and more reliable than a power relay.

A very different way to obtain the functions of a line conditioner is shown in Fig. 12-3a. Isolation is obtained by having an electric motor turn the shaft of a generator: there is no electrical connection between input and output sides of this circuit. The shaft between the motor and generator can be made of insulating material to provide complete isolation. Attenuation of high-frequency differential-mode noise is provided by the large moment of inertia of the rotating machinery: the rate of rotation of the generator can not change rapidly. An additional advantage of this arrangement is that interruptions of mains voltage, up to several hundred milliseconds in duration, do not affect the output voltage, since the rotating parts of the generator, motor, and shaft acts as a flywheel and stores mechanical energy.

The autotransformer is another type of AC voltage regulator. The autotransformer (also known as a "variable ratio transformer" or by the trade name "Variac") provides no isolation since the primary and secondary share turns on the same coil. The autotransformer fails to meet our requirements for a line conditioner and will not be considered further.

We have just reviewed the operation of AC regulators, one of the

GENERATOR

OUTPUT

SHAFT CAN BE
FIBERGLASS OR PLASTIC

AC
MOTOR

TRANSIENT

LINE

(a)

GENERATOR

OUTPUT

SHAFT CAN BE
FIBERGLASS OR PLASTIC

DC
MOTOR

AC TO DC
BATTERY
CHARGER

TRANSIENT

LINE

(b)
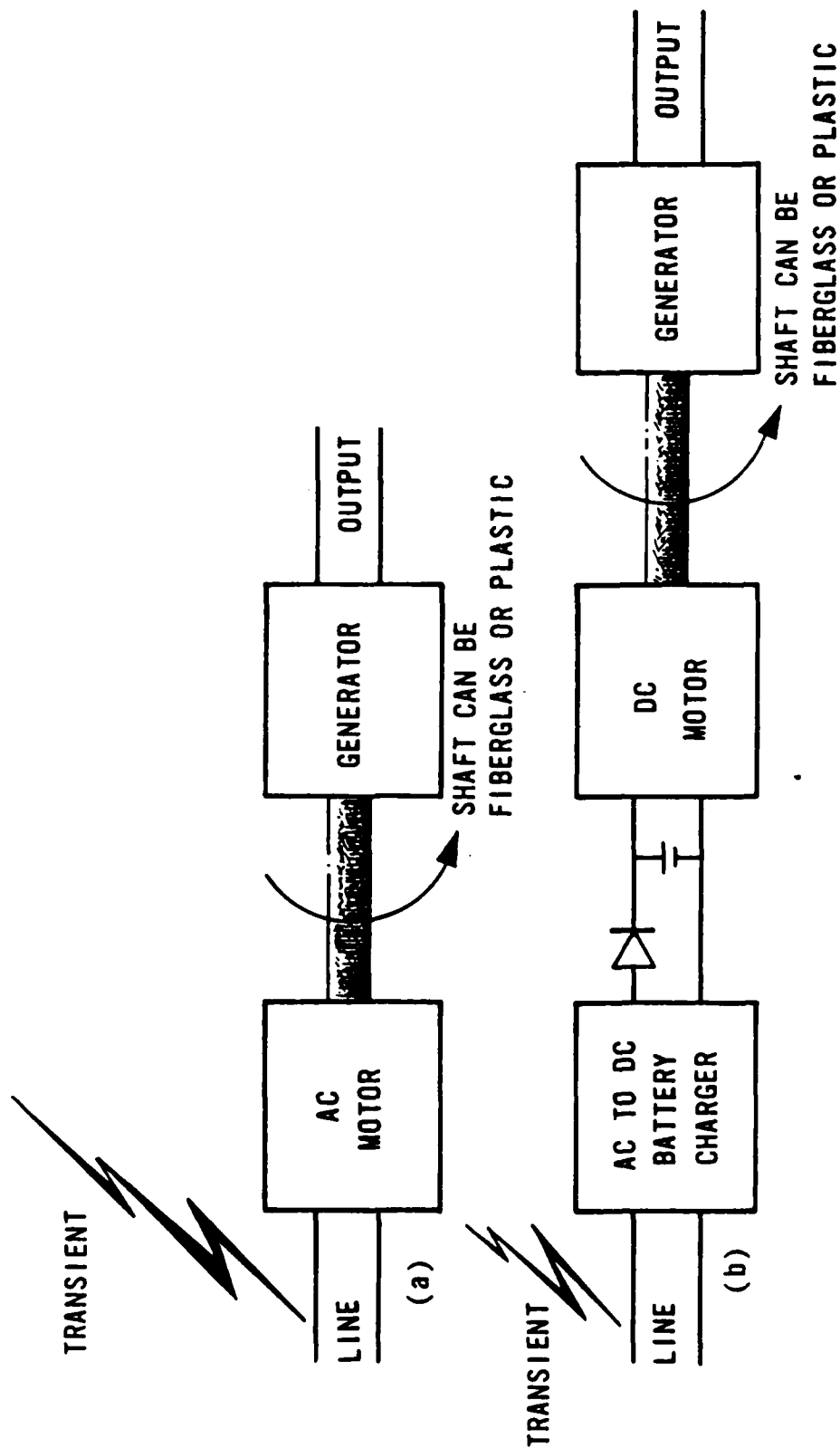
Fig.12-3

requirements for a line conditioner. The second requirement of a line conditioner is that it provide differential-mode attenuation for high frequency noise. This can be accomplished by including some inductance in series with the transformer primary, or by placing an inductor-capacitor filter module in series with the input side. If a relatively large series inductor is inserted in the primary side, the voltage across both primary and secondary coils of the transformer will be reduced. We can compensate for this effect by increasing the number of turns in the secondary coil. The large series inductance in series with the primary coil gives much better rejection of noise at frequencies below 0.5 MHz than common inductor-capacitor filter modules.

Line conditioners are manufactured in the U.S.A. by the following companies, which are listed in alphabetical order. The reader is advised that this list is not complete, but probably does contain most of the larger companies.

> Elgar, division of McGraw-Edison
>
> General Electric
>
> Gould, Power Conversion Div. (formerly Deltec)
>
> Frequency Technology (Isoreg^{TM})
>
> Sola, division of General Signal
>
> Topaz

The line conditioner is designed to solve most of the problems encountered with poor quality AC power. However, it does not solve two problems: (1) temporary loss of AC power, and (2) protection against severe

transient overvoltages.

Uninterruptible Power Supplies

The first of these problems can be solved by obtaining an
"uninterruptible power supply" (commonly called a "UPS"). One type of UPS,
which is shown in Fig. 12-4, contains batteries and an inverter (a circuit to
convert DC to AC). When the input AC power is lost (or the rms input voltage
is less than about 105 volts), the batteries and inverter provide continuous
AC power at the output terminals of the UPS. A small UPS can typically
operate its full load (e.g. 400 VA) between 10 and 30 minutes before the
batteries are completely discharged. When the input AC power is restored, the
batteries are recharged.

A UPS is particularly nice to have for a computer system, so that the
system does not "crash" during brief AC power outages. If the input power is
to be interrupted for longer than the UPS can operate the load, then the
computer system can be shut-down in an orderly way with all data or
information preserved on magnetic disks or magnetic tape. We note that large
computer systems require cooling, usually air conditioning, to prevent thermal
damage to the electronics. It is usually uneconomical to purchase a UPS that
has adequate capacity to operate the cooling system and computer. Since the
computer should be operated for less than 20 minutes without cooling, this
places a limit on the maximum practical size of the UPS.

If operation of critical loads during AC power interruptions of more than
about 30 minutes is desirable, a diesel engine and a generator may be more

TO LOAD

AC OUT

INVERTER

CONTROL

BATTERY CHARGER

AC MAINS
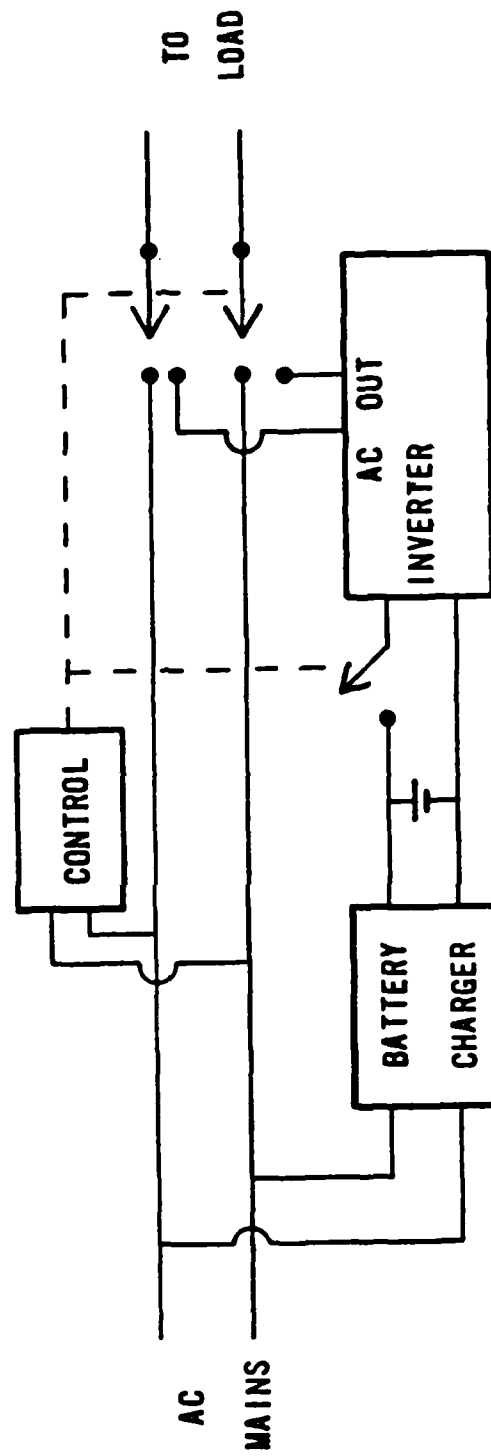
Fig.12-4

economical than a UPS that is operated from lead/acid batteries. This is particularly true for large systems that include air conditioners.

The motor/generator isolation unit can be modified, as shown in Fig. 12-3b, to be an uninterrruptible power supply.

How is a UPS to be connected with a line conditioner? There are two possible ways, which are shown in Fig. 12-5. The surge protection module, which will be discussed in detail in the next section, removes severe transient overvoltages from the mains. First, we discuss when the arrangement of Fig. 12-5a is preferable.

If the UPS has a non-sinusoidal output waveform (square wave inverters are common in less expensive UPS models), then a ferro-resonant line conditioner could be placed downstream from that UPS, as shown in Fig. 12-5a, to provide a more nearly sinusoidal waveform to the load.

If the output of the UPS is connected by a relay to either (1) the input of the UPS or (2) the output of the inverter, typically 5 to 10 milliseconds is required for the relay to change state. During this time no power is delivered to the output terminals of the UPS. By using a ferro-resonant line conditioner downstream from the UPS, as shown in Fig. 12-5a, we obtain more nearly continuous power to the load, since the resonant output circuit continues to oscillate for a few cycles in the absence of input power. However, such a situation will draw a large initial surge current from the UPS and may cause the circuit breaker at the UPS output to trip.
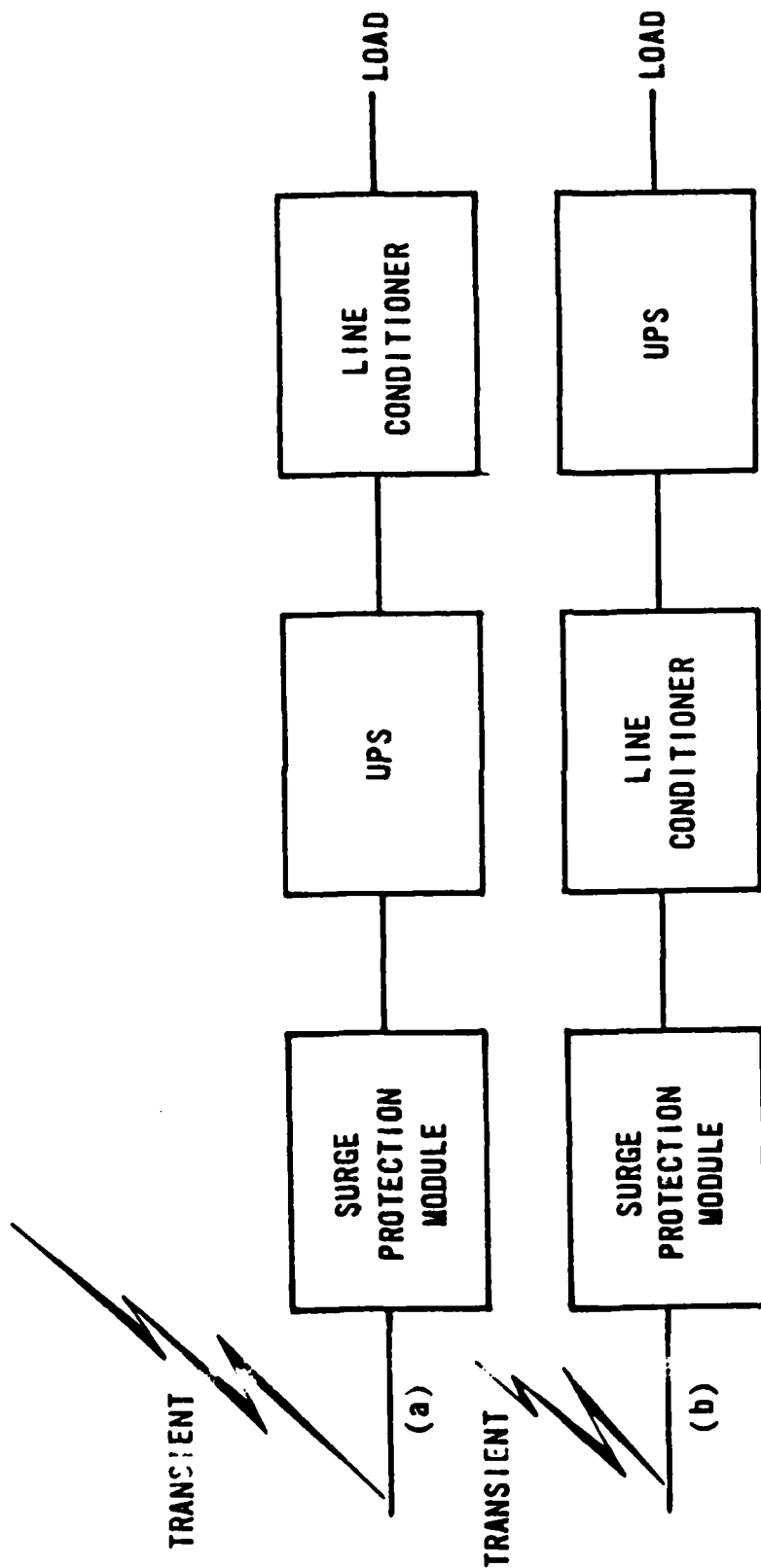
Fig.12-5

When the rms voltage at the load must be tightly regulated, a ferro-resonant line conditioner downstream from the UPS, as shown in Fig. 12-5a, is probably a good idea. A ferro-resonant transformer could be expected to have better output voltage regulation than a typical UPS. However, the user should check the specifications of the particular models that he intends to use to be certain.

The circuit of Fig. 12-5b is preferable when the line voltage if often between 95 and 110 volts rms. The line conditioner will boost these low voltages to acceptable levels and prevent the UPS from draining its batteries.

Surge Protection Modules

The other problem that is left unsolved by a line conditioner alone is that of severe transient overvoltages. This problem is solved by obtaining a surge protection module. These modules usually contain metal oxide varistor(s), possibly with the addition of an inductor-capacitor filter module.

Fig. 12-6 shows the schematic diagram of a comprehensive module which the author has designed and used. The parts list is given in Table 12-I.

The varistors V_1, V_2 are included to attenuate common-mode transient overvoltages. These devices should remove most of the overstress.

The capacitor C_1 is connected directly between the non-grounded terminals of V_1 and V_2 with minimal lead length. The purpose of this capacitor is to

COMPREHENSIVE AC POWER LINE TRANSIENT PROTECTION CIRCUIT
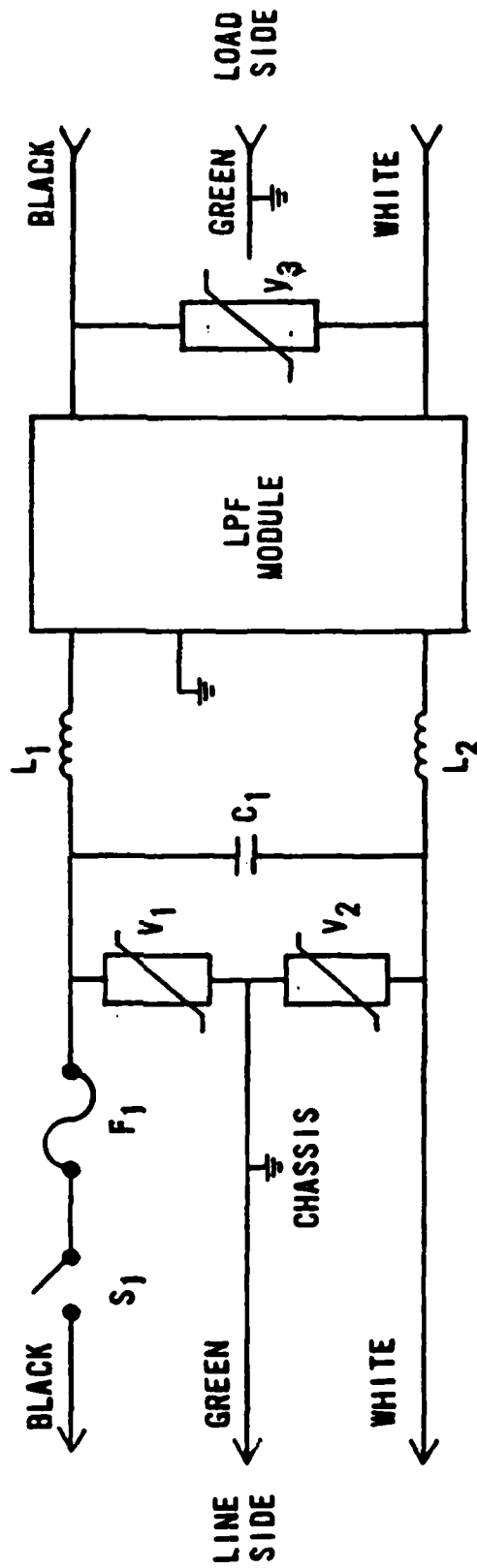
Fig. 12-6

reduce the rate of rise of differential-mode transients. It may not be necessary. Bell (1975, p.108) recommends at least a 600 volt DC rating when C_1 is to be used on 120 volt rms mains.

Inductors L_1 and L_2 are the point-to-point wiring between varistors V_1 and V_2 and the line side of the low-pass filter module (or to V_3 if the LPF is not installed). By placing many turns in these wires, we create more inductance in this path than in the shunt path through V_1 or V_2. Therefore, fast transient overvoltges with large amplitude are forced through V_1 or V_2 to ground. These inductors are an important part of the circuit.

The low pass filter module (LPF) is included to prevent radio frequency interference from perturbing the operation of the load. If interference above 50 MHz is likely, lossy line should be included between varistors V_1, V_2 and the LPF module. This is suggested because lumped-element filters do not work well at very high frequencies. In this situation, the output of the LPF module should be on the opposite side of a shielded enclosure from the line side of the LPF module.

Varistor V_3 is the final transient protection element. It is included to clamp the differential-mode output voltage to an acceptably low level (about 300 volts).

The fuse and switch are not necessary for transient protection, and are provided for convenience. In order for the fuse to be a credible safety device it should be rated as noted in the parts list.

Substantial, but less comprehensive, transient protection can be obtained by a much simpler circuit: one can install a varistor inside a common outlet strip to provide differential-mode protection. This varistor should have specifictions similiar to V_3 in the surge protection module that was discussed above. The varistor should be connected between the white and black wires to give differential-mode protection. If possible, two additional varisotrs can be included: one between the white and green pair, the other between the black and green pair. Each varistor should, of course, have minimal length of leads to avoid extra parasitic inductance. Installation of the varistor is more convenient in outlet strips that have solid non-insulated copper bus wire inside for the AC power and ground, than models with insulated wire inside.

We mention that all of the units in a system must be conencted to a line conditioner and surge protection modules, in order that the system be protected from transient overvoltages. This simple rule is often violated in computer systems, where the computer is connected to the mains through a line conditioner and surge protection modules, but not the peripherals such as the printer or terminals. Since the peripherals are connected to the computer via the communications interfaces, transient overvoltages could propagate from the mains through the peripherals to the computer. The simplest solution is to have one large line conditioner that serves all of the computer system. However, if some of the peripherals are to be located more than about 4 metres from the computer, separate line conditioners for the remote periphrals may be advisable.

VARISTOR IN CHASSIS

In a chassis that consumes less than 500 watts of AC power from a 110 to 120 V_{rms} mains, and does not contain a potent source of transients, an appropriate metal oxide varistor might have the following specifications.

1. $185 < V_N < 230$ volts   (V_N is the varistor voltage at 1 mA DC current)

2. less than 350 volts across the varistor during a 8×20 μs test current with a 100 ampere peak value

3. will survive a 4 kA peak current (8×20 μs waveshape) without rupturing case or loss of protective function

4. able to tolerate a at least a million pulses of 50 ampere peak current (8×20 μs waveshape) without changing V_N by more than ±10%.

In addition, the chassis must contain appropriate RFI/EMI filters in series with the mains to meet requirements for radio frequency interference. It would be desirable to put the varistor between the line cord and the filter. However, the filter is usually located at the socket for the line cord, so in modifying an existing chassis, the varistor must be located between the filter and the remainder of the electronics in the chassis.

TRANSIENT SUPPRESSION AT SOURCE

It can not be too strongly emphazised that that transients due to switching reactive loads should be suppressed at their source, rather than allowed to pollute a system. Relays, solenoids, and motors should have a

metal oxide varistor or bipolar avalanche diode installed across the coil (and across the contacts of a relay) as shown in Fig. 12-7a to suppress transients. If the relay, solenoid, or motor operates from DC (no polarity reversals during normal operation), then a rectifier that is connected across the coil, as shown in Fig. 12-7b, offers the best clamping. A system that is protected in this way will be a "good neighbor" and is less likely to cause problems when it is installed near a sensitive system.

If it is not feasible to suppress transients at their source, then a system that produces transients must be isolated from the mains by a surge protection module. Common sources of transients are large motors (e.g. air conditioners, refrigerators, elevators, etc.), fluorescent lamps, neon signs, photocopiers, and vending machines.

(a)

(b)
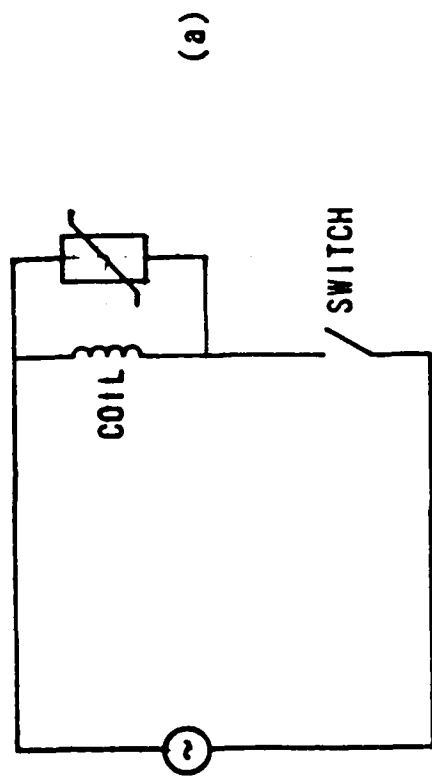
Fig.12-7
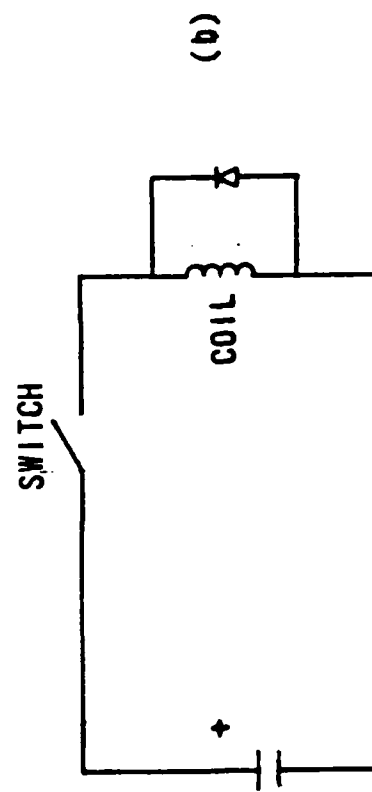
Table 12-I
PARTS LIST FOR SURGE PROTECTION MODULE

V_1, V_2 metal oxide varistors

(e.g. General Electric V130PA20A or V150PA20A)

specifications: $185 < V_N < 285$ volts   (at 1 mA DC)

         $V < 420$ volts at $I = 100$ A   (8×20 µs wave)

         $V < 800$ volts at $I = 5000$ A   (8×20 µs wave)

able to tolerate at least a million pulses of 60 ampere peak current

(8×20 µs waveshape) without changing V_N by more than $\pm 10\%$

will survive a 6.5 kA peak current (8×20 µs waveshape) without

rupturing case or loss of protective function

V_3 metal oxide varistor

(e.g. General Electric V130LA20A or V130LA20B)

specifications: $185 < V_N < 230$ volts   (at 1 mA DC)

$V < 340$ volts at $I = 100$ A   (8×20 µs wave)

same pulse lifetime and peak current rating as V_1 above

C_1 (optional) capacitor, 0.01 µF $\leq C \leq 0.22$ µF

DC rating ≥ 1000 V, must be "self-healing"

L_1, L_2 about ten turns of insulated 14 or 16 AWG copper wire

(internal diameter of coil is about 1 cm)

LPF   (optional) commercial mains low-pass filter module

This filter should withstand at least a 1.45 kV DC test between the

two line terminals, or between either line terminal and ground, for

at least one minute. It is also desirable that the filter provide

both common-mode and differential-mode insertion loss of at least 30

dB between 0.5 MHz and 30 MHz (in a 50 Ω test fixture).

F_1 (optional) fuse

Specifications: current rating equal to or less than that of low-

pass filter module, but not greater than 20 amperes. This fuse

should be rated to interrupt 10 kA on a 250 volt rms mains.

PROTECTION OF DC POWER SUPPLIES

INTRODUCTION

Transient protection of DC power supplies is essential because loads can be damaged by excessive power supply voltages. U.S. Military Handbook 419 (1982, p.1-84) states that "power supplies (5 to 48 V) operating from AC inputs and supplying operating power for solid-state equipment always require internal transient protection [emphasis added]." Fig. 13-1 shows a DC power supply and several critical loads. There are four different situations, which are illustrated in Fig. 13-1, where the loads are vulnerable to damage.

1. Transient overvoltages can enter the system through the mains, propagate through the DC power supply module, and affect all of the loads.

2. Transient overvoltages can enter the system on a data line, propagate through an amplifier onto the power supply line, and then affect other loads.

3. A load that requires a DC power supply current with a large time rate of change (owing to both high frequency and large amplitude) will produce a voltage fluctuation on a DC power bus. This voltage fluctuation is caused by parasitic inductance in the DC power supply bus and large values of dI/dt, where I is the power supply current to the load. Notice that a large value of di_3/dt in Fig. 13-1 will also affect the power supply voltages at the other two loads.

4. Failure of the voltage regulator in the DC power supply module can produce a sustained overvoltage condition that can destroy the loads.

The effect of item 3, voltage fluctuations on the DC power supply bus, is usually not associated with damage to the loads, however it can produce

OVERSTRESS

OVERSTRESS

DATA LINE

MAINS

AC
IN

DC
OUT

POWER SUPPLY
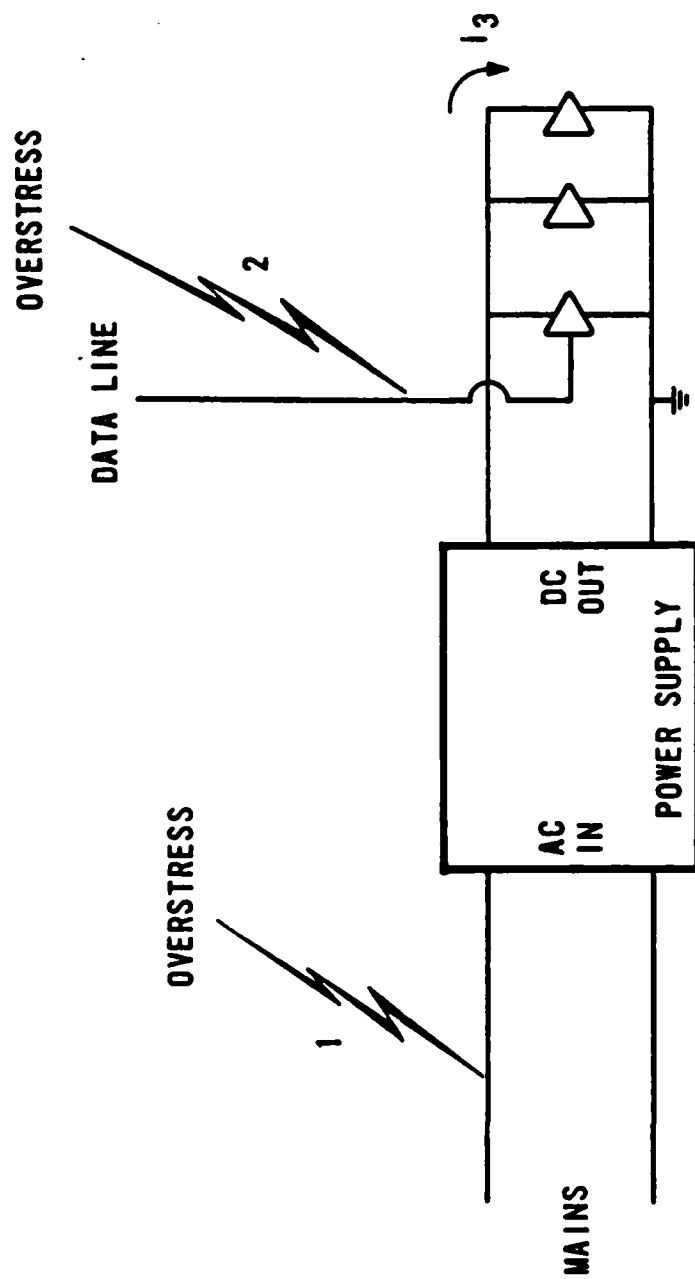
Fig.13-1

excessive noise and upset. We mention it here becasue it is part of the
general electromagnetic compatibility problem and because techniques for
dealing with it also attentuate destructive overvoltages on the DC supply
line. Voltage fluctuations owing to parasitic inductance in the power supply
conductors is usually reduced by installing bypass capacitors near each load
that has a large time rate of change of DC supply current, and by distributing
bypass capacitors along the DC power supply conductors. This was discussed in
Chapter 10 on parasitic inductance. We will also discuss it below.

A simple, unprotected, DC power supply circuit is shown in Fig. 13-2.
For convenience, we shall only discuss a circuit with a positive output
voltage. The transformer reduces the mains voltage to a lower value in an
energy-efficient way. Typical secondary voltages are 20 volts rms for 15 volt
DC supplies, 10 volts rms for 5 volt DC supplies. The four rectifiers, D_1 -
D_4, convert the sinusoidal secondary voltage into a waveform that always has
the same polarity. The filter capacitor, C, provides a quasi-DC voltage by
storing charge when the output voltage of the rectifiers is greater than the
capacitor voltage and by releasing charge when the output voltage of the
rectifiers is less than the capacitor voltage. The voltage regulator module
provides a constant output voltage by varying the effective series resistance
between the input and output terminals. The regulator module forms a voltage
divider with the load resistance. By using negative feedback, the regulator
maintains a constant output voltage for a wide range of values of load
resistance (or load current) and input voltage.

We will limit the discussion to integrated circuit voltage regulators,
since these are inexpensive and deliver good performance. In order to

MAINS

INTEGRATED CIRCUIT
VOLTAGE REGULATOR

IN   OUT

COM

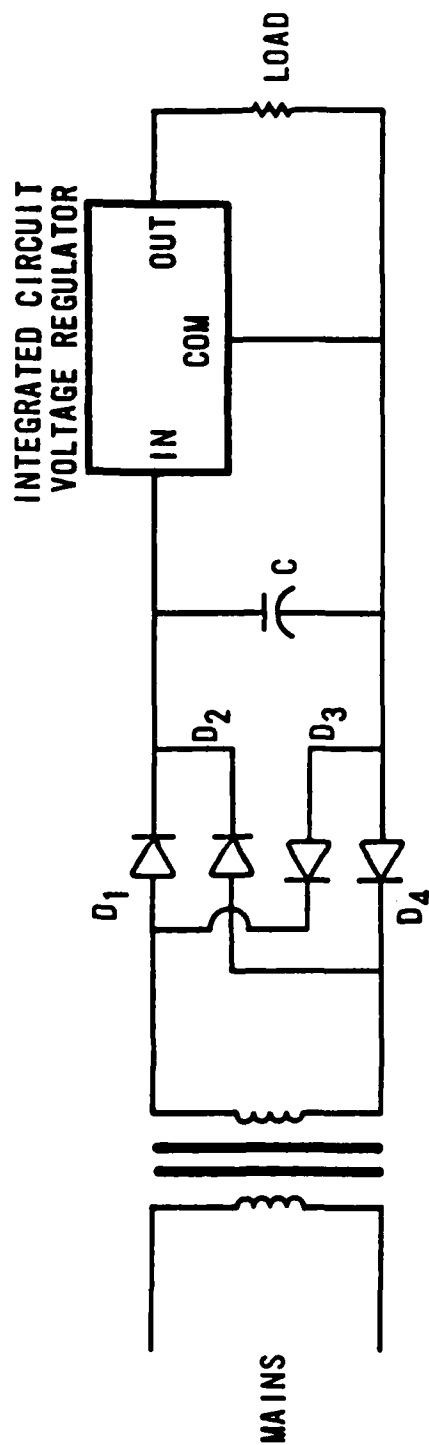D_1 D_2

D_3 D_4

C

LOAD

Fig. 13-2

provide good performance, the value of $(V_{in} - V_{out})$ should be greater than about 3 volts. However, larger values of $(V_{in} - V_{out})$ waste energy and contribute to thermal stress on the voltage regulator components, provided that the load current is also large. If V_{in} is too large, the regulator module may be destroyed.

There are four different groups of components that need protection.

1. The transformer needs protection from excessive primary voltage, both common-mode and differential mode.

2. Rectifiers need protection from excessive current and excessive reverse voltage.

3. The input port of the voltage regulator and electrolytic filter capacitor needs protection from excessive voltage.

4. Output port of the voltage regulator and the loads that are connected to it require protection from overvoltage.

We now discuss how to protect each of these groups of components.

PROTECTION FROM TRANSIENTS ON MAINS

A transformer is quite robust, when compared to the other components in a DC power supply. Nevertheless, we must protect the transformer because its failure will be catastrophic for the power supply. If the transformer fails with a short-circuit between primary and secondary, the full mains voltage can be applied to the rectifers, filter capacitor, and voltage regulator with probable destruction of all of them. If the transformer fails with an open-circuit in either the primary or secondary winding, the DC power supply will cease to function.

In order to obtain good isolation from the mains, the transformer should have either a properly grounded electrostatic shield between the primary and secondary coils, or the two coils should be non-concentrically wound. These specifications reduce the coupling capacitance between the primary and secondary coils. In addition, it would be desirable for the transformer to be specified to survive 2500 volts rms continuously between primary and secondary windings with no degredation in insulation. These requirements assure that the transformer is robust.

Metal oxide varistors should be connected from each side of the mains to ground to provide protection from excessive common-mode voltage at the primary winding of the transformer. A third metal oxide varistor should be connected across the primary winding of the transformer to provide protection form excessive diffferential-mode voltage. This is shown in Fig. 13-3, and described in detail in Chapter 12 on mains applications. In years past it was considered good form to use capacitors in the locations shown by V_1, V_2, and V_3. Common units were ceramic disks with a capacitance between 0.005 µF and 0.02 µF and a working voltage rating of 1.4 kV DC. The metal oxide varistors are clearly superior, owing to their capability to clamp at a small voltage even when a large surge current (with a large charge transfer) is present. Moreover, the 20 mm disk metal oxide varistors with V_N of about 200 volts have a parasitic capacitance of about 0.002 µF. Thus the use of the metal oxide varistors automatically includes the capacitative shunt.

These metal oxide varistors protect the rectifiers, as well as the transformers, from excessive voltages. For example, consider the situation
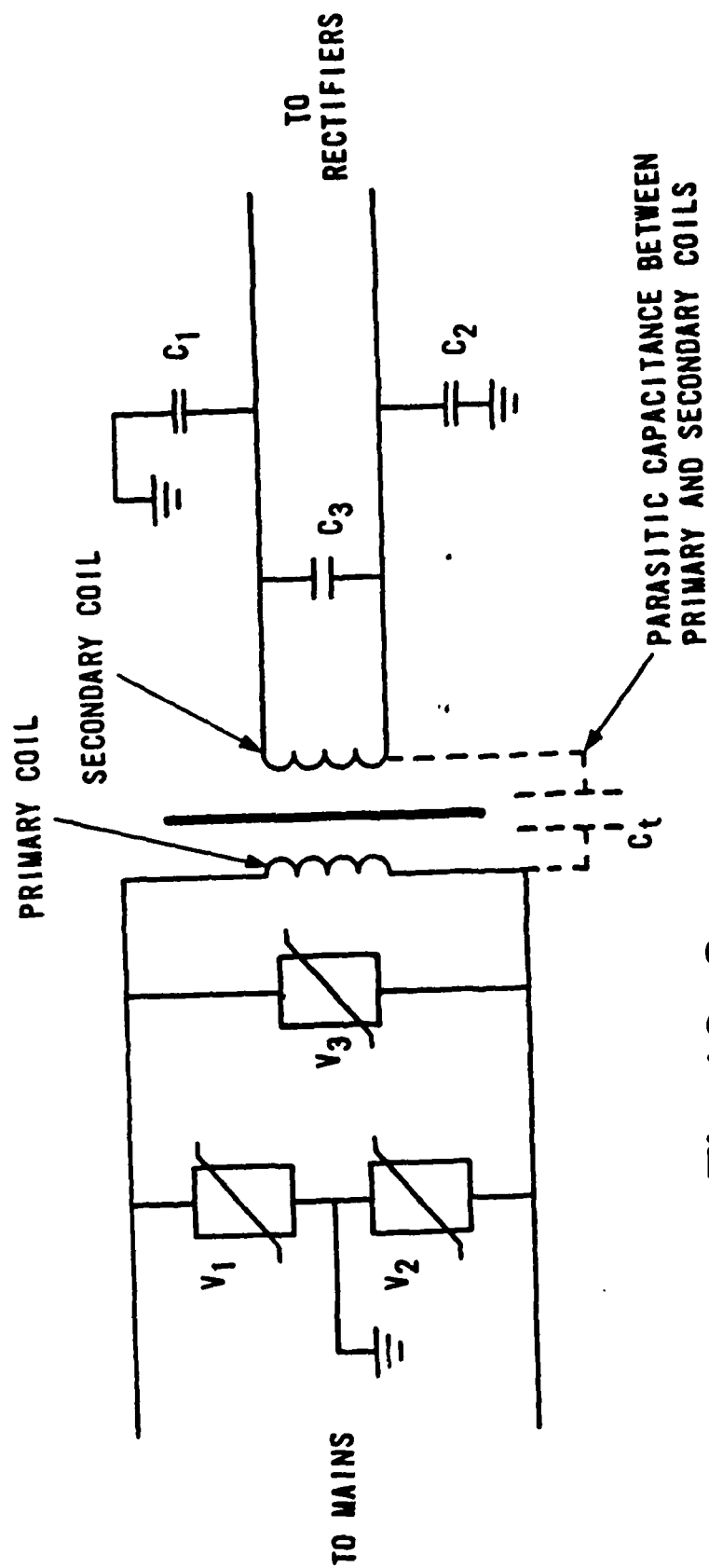
Fig. 13-3

where the primary is connected to 120 volt rms and the secondary voltage is 28 volts rms, a ratio of 1:0.23. If the varistor limits the peak differential-mode primary voltage to 300 volts, then the peak secondary voltage would be expected to be about 70 volts. This should be well within the reverse breakdown rating of the rectifiers, which typically have a reverse breakdown rating of 400 volts. The capacitor C_3, which is shown in Fig. 13-3, provides a shunt path for differential-mode transient currents. A typical value of C_3 is between 0.01 µF and 0.1 µF. This capacitor must be mounted near the transformer secondary terminals to minimize electromagnetic radiation.

We now consider common-mode attentuation over overvoltages by the transformer, capacitors C_1 and C_2, and varistors. The two capacitors C_1 and C_2, which are shown in Fig. 13-3, constitute a voltage divider with the parasitic capacitance between primary and secondary coils, C_t. This voltage divider attenuates common-mode transients, in the following ratio.

$$(v_{out}/v_{in}) = C_t/(C_1 + C_t)$$

Typical values of C_t are less than 1 pF if a grounded electrostatic shield is used between the primary and secondary, and about 30 pF if the two coils are non-concentrically wound. Typical values of C_1 and C_2 are about 0.01 µF each. These typical values give a (v_{out}/v_{in}) ratio of of less than 10^{-4} when an electrostatic shield is used, and about 1.5×10^{-3} for non-concentrically wound transformers. If varistors V_1 and V_2 limit the maximum value of V_{in} to 300 volts, the common-mode output voltage will be less than 0.5 volt. This places negligible stress on the DC power supply, provided the transformer can withstand 300 volts between the primary and secondary. By also requiring a 2500 volt rms insulation rating, we obtain at least an order of magnitude safety factor.

The voltage rating of capacitors C_1, C_2, and C_3 should be at least three times the rms secondary voltage or 100 volts, whichever is greater.

Capacitors C_1, C_2, and C_3, along with varistors V_1, V_2, and V_3, should be installed with minimal lead length to reduce parasitic inductance in series with the shunt path.

Lasitter and Clark (1970, p.76) recommend placing a 1 µF non-electrolytic capacitor in the position shown as C_2 in Fig. 13-3. This relatively large capacitance, together with the leakage inductance of the transformer, forms a single-pole low-pass filter. However, the parasitic inductance of the 1 µF capacitor may limit the performance of this filter at high frequencies.

Additional protection for the rectifiers can be obtained by specifying their maximum reverse voltage rating (also called "peak inverse voltage") to be 4 times the normal rms secondary voltage or 400 volts, whichever is the greater.

Additional protection for the filter capacitor, C in Fig. 13-2, can be obtained by specifying a DC working voltage that is at least 2.5 times the normal rms secondary voltage (or at least 1.7 times the peak secondary voltage). One should remember that the secondary voltage is greater than the nominal value (which is measured at the maximum rated secondary current) when the transformer has a small secondary current, the no-load condition for the power supply. There is no point in allowing the no-load condition to stress the filter capacitor.

BYPASSING

Sometimes one encounters a unregulated or quasi-regulated DC power supply bus that serves one or more distant loads. A voltage regulator should be located near each load. The input terminals of these voltage regulators should have several bypass capacitors in parallel that includes large electrolytic capacitors (e.g. 100 µF) as well as smaller electrolytic capacitors (e.g. 1 µF to 10 µF) and ceramic capacitors (e.g. 0.01 µF to 0.1 µF). The parallel combination of capacitors is necessary to counter the effect of the parasitic inductance in the long cable between the DC source and the voltage regulator. Fig. 13-4 shows the magnitude of impedance vs. frequency for the parallel combination of the following four capacitors:

	C	L	series resistance
aluminum electrolytic	100 µF	280 nH	0.5 Ω
tantalum electrolytic	6.8 µF	5.8 nH	0.2 Ω
ceramic disc	0.1 µF	7 nH	0.2 Ω
ceramic CK06 style	0.01 µF	2.8 nH	0.2 Ω

The 0.1 µF ceramic disk capacitor is not essential, but it does reduce the impedance between 10 MHz and 20 MHz by about a factor of two. A 100 µF aluminum electrolytic capacitor is suggested instead of a 100 µF tantalum electrolytic capacitor since the tantalum unit is much more expensive and the superior electrical properties of tantalum electrolytics are not important in this application.

A voltage regulator should always be located near the load, and usually be within about 30 cm if possible. If the voltage regulator is far from the
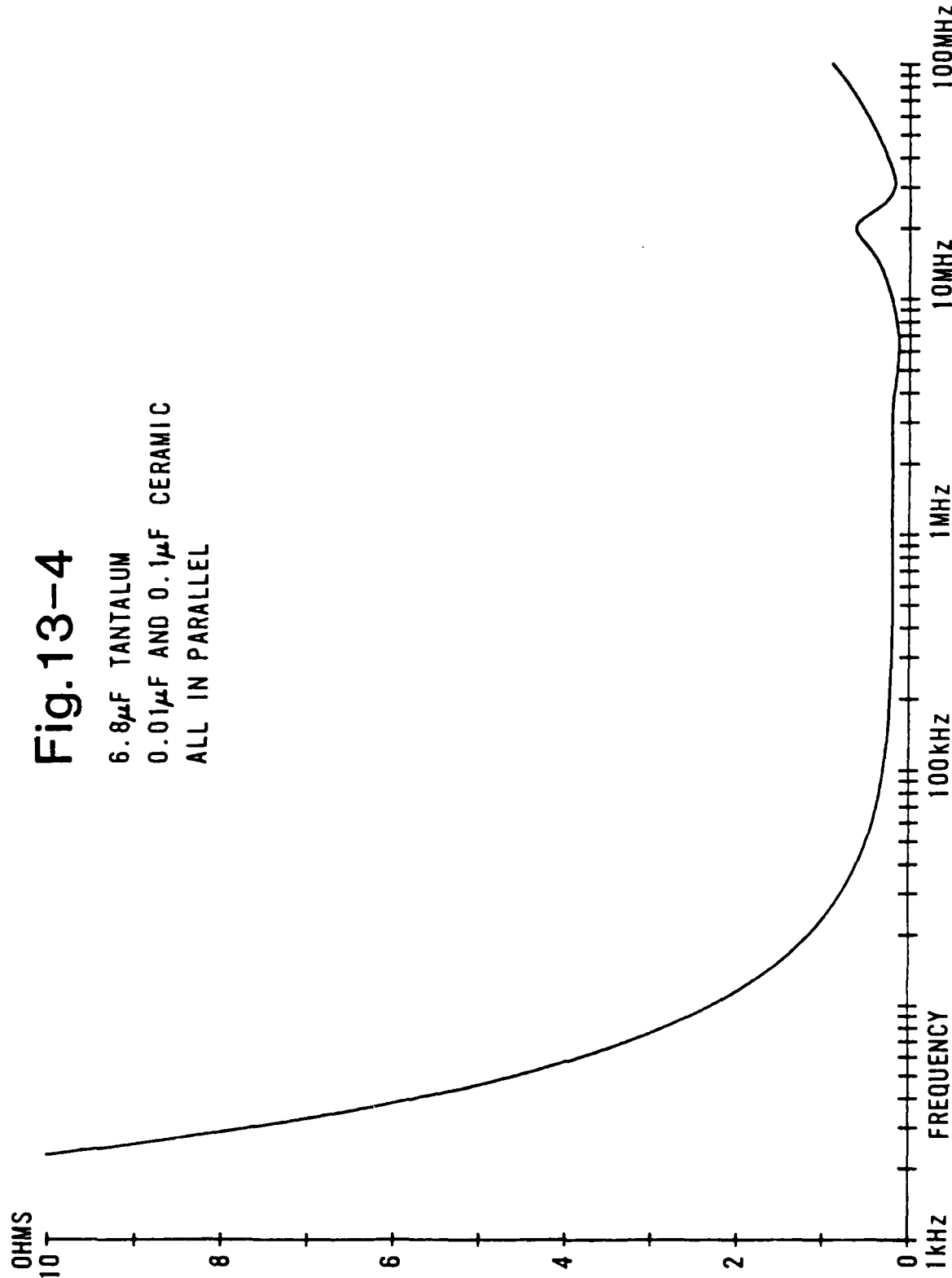
Fig. 13-4

6.8μF TANTALUM
0.01μF AND 0.1μF CERAMIC
ALL IN PARALLEL

load, the regulator can not be expected to maintain a constant load voltage, owing to parasitic inductance and resistance of the wiring.

PROTECTION OF REGULATORS

A positive integrated circuit voltage regulator can be damaged by excessive input voltage, or by a value of $(V_{in} - V_{out})$ that is more negative than -0.6 volts. We shall now consider how to protect against these two damaging conditons.

Typical maximum values of V_{in} are about 35 volts for most integrated circuit regulators, although some regulators with V_{out} = 5 volts have a maximum input voltage as small as 20 volts. When protection is necessary, the best way to protect against excessive input voltage is to use an avalanche diode as shown in Fig. 13-5. The avalanche diode is the agent of choice owing to its highly non-linear relation between current and voltage. The breakdown voltage of the diode should be chosen to be slightly less than the maximum tolerable input voltage of the regulator circuit. However, the breakdown voltage of the diode must be greater than the peak input voltage during normal operation of the DC power supply. If the normal operation of the DC power supply should cause the avalanche diode to conduct, thermal destruction of the avalanche diode is almost certain. Owing to this hazard, it would be preferable to design power supplies that limit the input voltage to the regulator to safe values by using only the components shown in Fig. 13-3. If the varistors on the mains and the capacitors on the secondary side are inadequate to assure the survival of the regulator during proof testing, an avalanche diode can be added as a last resource.
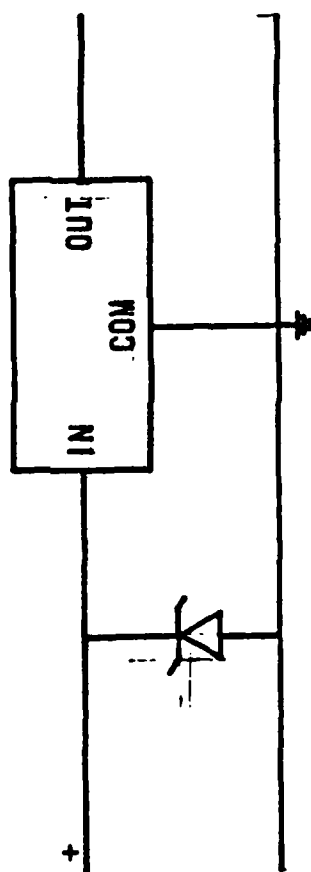
Fig.13-5

The avalanche diode, as shown in Fig. 13-5, is recommended if a long cable is present between the source (transformer) and the voltage regulator. Transients that enter the system on this cable will not be attenuated by the varistors on the mains, or by the capacitative voltage division, as discussed above in connection with Fig. 13-3. Therefore, the avalanche diode is a reasonable choice in this situation.

When one is tempted to use an avalanche diode to protect a voltage regulator, it is recommended that one also consider using a voltage regulator with a large maximum input voltage (e.g. model LM317HV for positive voltages, LM337HV for negative voltages). The "extra" margin of tolerable input voltages will allow an increased safety margin between the normal operating voltages and the avalanche diode conduction voltage.

The value of $(V_{in} - V_{out})$ must never be more negative than about -0.6 volts; violation of this rule will destroy an integrated circuit voltage regulator. Such a condition can arise if there is a short-circuit on the input side of the voltage regulator while the output voltage is maintained by bypass capacitors. Such a condition can also arise if a transient forces the voltage at the output terminal to a large positive value. Protection for the voltage regulator against these hazards can be obtained by use of the circuit in Fig. 13-6. The diode in Fig. 13-6 should be a rectifier that is capable of discharging the bypass and filter capacitors. This diode is normally reverse-biased by about 3 to 5 volts.

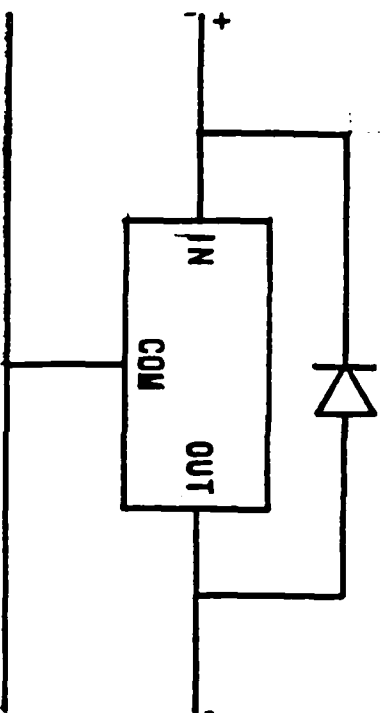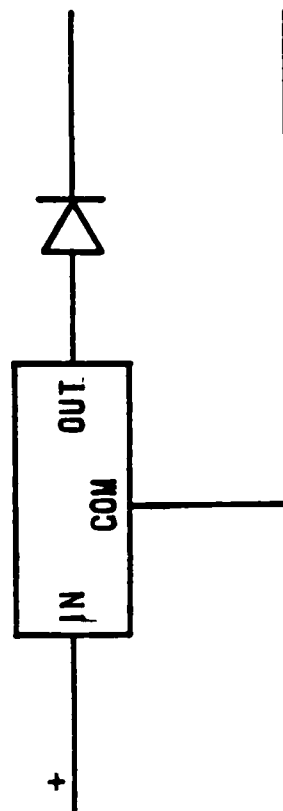The circuit of Fig. 13-7 also protects the positive voltage regulator

Fig.13-6

Fig. 13-7

from positive current passing into the output terminal. However, using a
series diode, as shown in Fig. 13-7, has several disadvantages. First, the
diode degrades the voltage regulation during normal circuit operation. The
voltage across this diode can vary between 0.6 and 1.0 volt, depending on the
current. Second, the diode may offer little protection from fast risetime
transients. Since the diode is normally conducting, and an overvoltage causes
the diode to be reverse-biased, the diode's response time is on the order of
the reverse recovery time. The reverse recovery time is typically on the
order of microseconds for most rectifier diodes.

PROTECTION FROM TRANSIENTS AT LOADS

The load can be protected against transients that are introduced on the
output side of the voltage regulator by the circuit shown in Fig. 13-8a. The
minimum value of the avalanche diode breakdown voltage should be chosen to be
about 1.2 times the maximum output voltage of the regulator circuit. The
words "minimum" and "maximum" should be understood to include effects due to
device tolerance, temperature, changes in load current, etc. This should
preclude the avalanche diode conducting during normal operation of the system.
For example, a 6.8 volt ± 5 % diode could be used to protect a 5 ± 0.3 volt
power supply bus. In this example there would be at least a 1.1 volt margin
between the maximum supply voltage and the minimum avalanche voltage.

U.S. Military Handbook 419 (1982, p.1-89) recommends that the minimum
avalanche diode voltage at a 100 microampere diode current be 1.05 times the
maximum output voltage of the regulator. This is a different way of
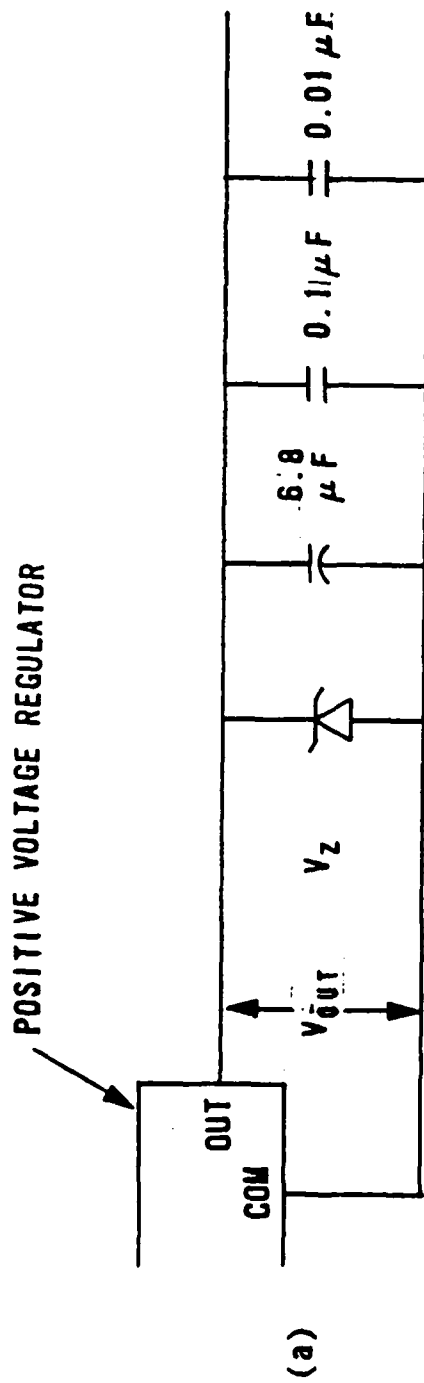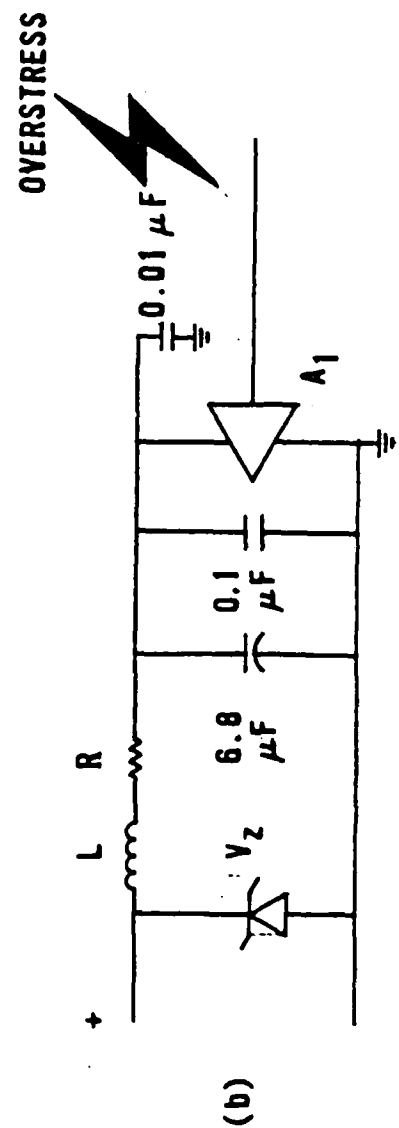expressing the same concerns as given in the previous paragraph.

POSITIVE VOLTAGE REGULATOR

OUT
COM

V_{OUT}

V_Z

6.8 μF

0.1 μF

0.01 μF

(a)

Fig.13-8

OVERSTRESS

0.01 μF

A_1

L   R

V_Z 6.8 μF 0.1 μF

+

(b)

The bypass capacitor network shown in Fig. 13-8a is included to provide a stable DC voltage even when the loads are drawing rapidly changing currents from the power supply bus. In addition, they provide limited protection against transients.

Provided that there are no loads that are prone to severe transients (e.g. line drivers, line receivers), one clamping diode and one 6.8 µF tantalum electrolytic capacitor will serve to protect a single printed circuit board. The 0.1 µF and 0.01 µF ceramic capacitors should be distributed over the area of the printed circuit board for optimum suppression of transients. Devices that are possible or likely sites for severe transients should have an avalanche diode located near the device, as well as a bypass capacitor network. If the device merely has a large time rate of change of power supply current, a tantalum capacitor and a ceramic capacitor should be located near the device.

Notice that the avalanche diode shown in Fig. 13-8a protects the loads from positive overvoltages, as well as protecting the electrolytic bypass capacitors and electronic loads from voltage reversals on the power supply bus.

When it is anticipated that very severe transients will be introduced through the load, series resistance and inductance might be introduced to reduce the stress on the avalanche diode, as shown in Fig. 13-8b. However, the series element(s) decreases the protection that is provided for the integrated circuit, by increasing the voltage that appears across load A_1. If

we put an avalanche diode in parallel with the capacitors in Fig. 13-8b, we return to the sitution where we must worry about the survivability of the avalanche diode. The proper resolution of this situation is to put overstress protection on the input signal line, and thus prevent the severe transient from arriving at the input terminal, as well as the power supply terminals, of load A_1.

DC CROWBAR

There may be situations for which the avalanche diode of Fig. 13-8 provides inadequate protection of the DC supply bus. DeSouza (1967) described a comprehensive crowbar circuit, shown in Fig. 13-9, for protecting loads that were connected to a 28 volt power supply. This circuit is called a "crowbar" because it short-circuits the power bus as if a metal crowbar were dropped across the two conductors. Crowbar circuits have the advantage of being able to conduct large currents for sustained periods of time, unlike many other transient suppression circuits. Crowbar circuits have the disadvantage that they interrupt power to critical loads and may cause upset of the system.

The normally closed relay is included to interrupt "follow-current" in either the spark gap or SCR. The fuse, inductor L_1, and resistor R_1 prevent damage to the SCR by large surge currents. If the SCR switches to the conducting state when the gate current is 10 mA, the SCR will conduct at a line voltage of 40.7 volts since

$$(40.7 - 0.7) \text{ volts} / 4 \text{ k}\Omega = 10 \text{ mA}$$

Rectifier D_1 provides clamping of reverse voltages. This allows the use of a SCR with a low reverse breakdown voltage (e.g. 25 volts). The inductor L_1,
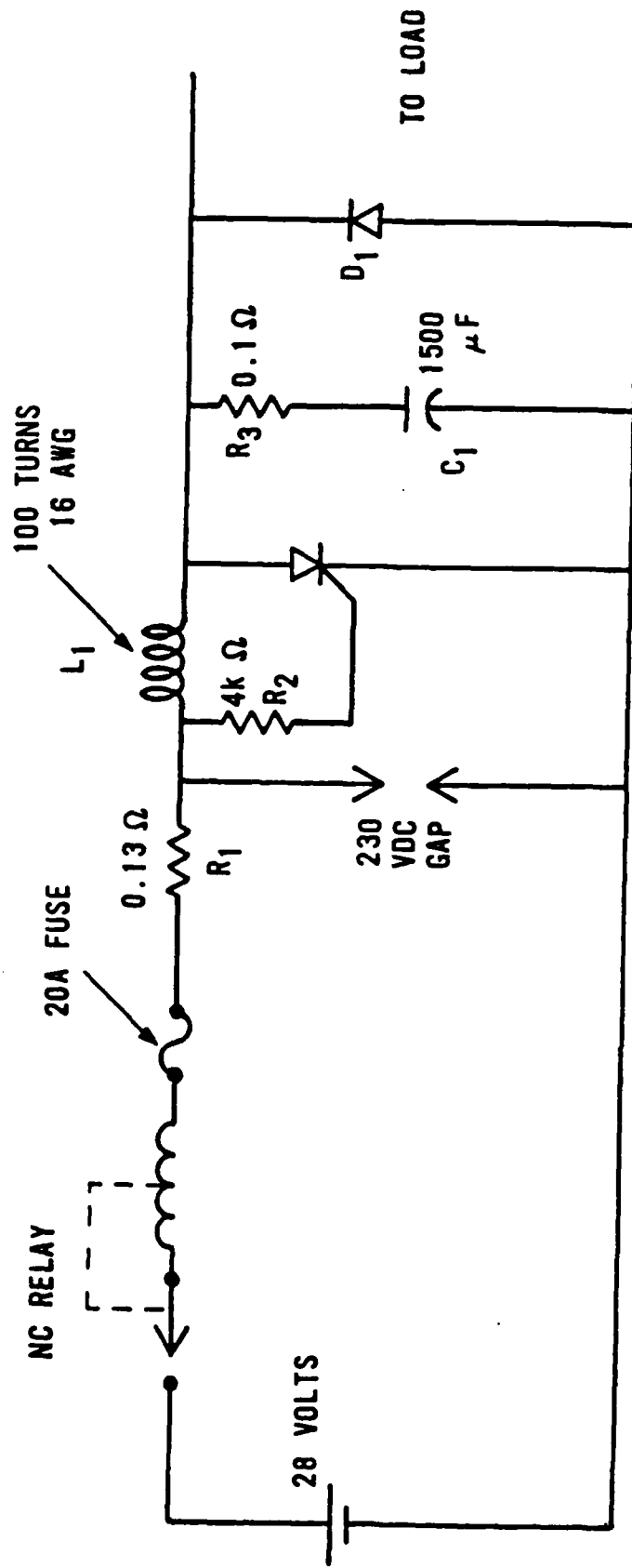
DE SOUZA'S DC POWER LINE PROTECTION CIRCUIT

Fig. 13-9

resistor R_3, and capacitor C_1 form an oscillation suppression network that attenuates transients caused by change of state of the relay, spark gap, SCR, or rectifier D_1.

DeSouza claimed that the use of a SCR allowed his circuit to have "a faster response time and a much greater current capacity than" circuits that use a zener diode. There is no debating that a SCR can conduct larger steady-state currents than an avalanche diode. If we compare a 10 watt SCR with a 10 watt, 40 volt avalanche diode, we find that the SCR can carry 5 amperes but the avalanche diode can carry only 0.25 amperes. This difference is due to the difference in conduction voltages: about 2 volts for the SCR and 40 volts for the avalanche diode. However, DeSouza is mistaken about the speed of response. If large currents flow through a SCR before the device is fully turned-on, a process which can take a few microseconds, the SCR can be damaged. In contrast, the avalanche diode turns on in much less than one nanosecond.

The circuit in Fig. 13-9 might be improved by including an avalanche diode in series with R_2 so that the cathode of the avalanche diode is connected to the positive line. The breakdown voltage of the avalanche diode could be selected to be slightly greater than the line voltage during normal operation (e.g. 33 V ± 5% for a 28 V line). This could permit the SCR to conduct during small transients that were nearer the normal operating voltage across the line. Alternately, one could specify an avalanche diode breakdown voltage that was slightly less than the maximum tolerable voltage across the load(s) in order to avoid nuisance crowbar actuations.

UNINTERRUPTIBLE POWER SUPPLY

If the mains power should be interrupted, the common DC power supply will be unable to continue to supply power to critical loads. A relatively inexpensive uninterruptible power supply is shown in Fig. 13-10.

This circuit uses N rechargable batteries connected in series to form a reserve power supply. The voltage regulator U_2 regulates the voltage across the batteries to a constant voltage across the critical loads. The voltage regulator U_1 regulates the battery-charging voltage. Diode D_1 prevents the batteries from discharging into the regulator U_1. The mains connection, transformer, and rectifiers have been omitted from Fig. 13-10 for simplicity.

Sealed, "maintenance-free," lead-acid cells are particularly desirable for the batteries because they (unlike nickel-cadmium batteries) can be connected indefinitely to a constant voltage source without degrading their life. The number of cells, N, is determined by the load voltage, V_{out}, the minimum desirable value of $(V_{in} - V_{out})$ for regulator U_2, and the "fully discharged" cell voltage. If we specify

$$(V_{in} - V_{out}) \geq 3 \text{ volts}$$

and use lead-acid batteries with a fully discharged potential of 1.6 volts/cell, we obtain Eqn. 1.

$$N \geq (V_{out} + 3 \text{ volts})/(1.6 \text{ volts}) \qquad (1)$$

The value of N, of course, must be an integer.

The value of output voltage from regulator U_1 is set by the values of R_1 and R_2. The value of R_1 is arbitrarily set to 220 Ω, and the value of R_2 is
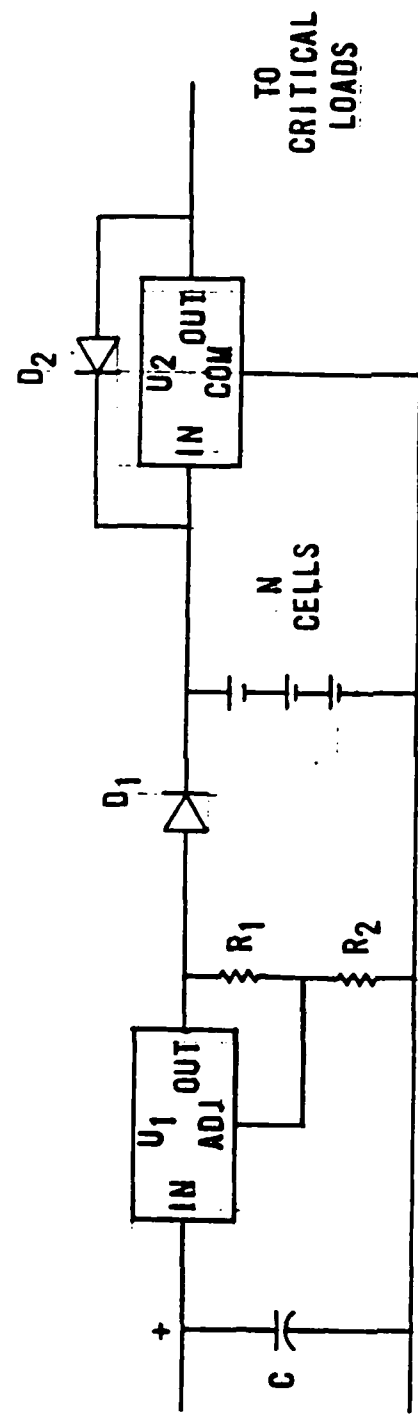
Fig.13-10

then calculated from Eqn. 2, where

$$V_1 = (V_{out1}/R_1)(R_1 + R_2) \qquad (2)$$

V_1 is the potential difference across the series connection of R_1 <u>and</u> R_2 and

V_{out1} is the potential difference between the output and adjustment terminals

of regulator U_1. We can neglect the current that flows in the adjustment

terminal of U_1 because it is much smaller than V_{out1}/R_1 when R_1 is 220 Ω. The

value of V_1 is determined by the proper charging voltage for N batteries in

series and the voltage drop across diode D_1. If the charging voltage per cell

is too small, the cells will take a long time to be partially charged and

never will be fully charged. If the charging voltage is too large, prolonged

connection to this voltage after the cells have been fully charged will damage

the cells. A reasonable value for the charging voltage per cell for lead-acid

batteries is between 2.25 and 2.40 volts. With this information we arrive at

Eqn. 3.

$$V_1 = (N \times 2.30) + 0.6 \text{ volts} \qquad (3)$$

Eqns. 2 and 3 can be solved for the value of R_2, given V_{out1}, N, and the

choice of 220 Ω for R_1.

The circuit of Fig. 13-10 has excellent voltage regulation, owing to the

two voltage regulators, U_1 and U_2, in series. The batteries also act like a

filter capacitor, since they have an incremental impedance, $\Delta V/\Delta I$, of the

order of a few milliohms.

The major disadvantage of the simple circuit in Fig. 13-10 is that the

regulator U_1 must be capable of supplying both the current to the loads,

through regulator U_2, and the battery charging current. When the batteries

are discharged and the mains are reconnected, the initial battery charging

current alone can be between 5 A and 10 A for 2.5 A-hr cells. Such a large
current can cause many popular three-terminal integrated circuit voltage
regulators to behave as constant-current sources owing to internal "short-
circuit protection." Alternately, the combination of the voltage drop across
the input and output terminals of U_1 and the large current can cause internal
thermal protection circuit to operate and decrease the output voltage below
normal values. While the internal short-circuit protection and thermal
protection circuits save regulator U_1, they may make it difficult to operate
the critical loads once the mains power has been restored.

The disadvantages of the simple circuit in Fig. 13-10 can be avoided by
using a separate power supply for the battery charging, as shown in Fig. 13-
11. The power transistor Q_1 and avalanche diode D_1 form a voltage regulator
for charging the batteries. The battery charging voltage is the breakdown
voltage of the avalanche diode minus the base-emitter voltage drop of Q_1.
Resistor R_1 limits the maximum charging current to a safe value for transistor
Q_1. Resistor R_2 limits the current in the avalanche diode D_1. Rectifier D_2
prevents the transformer secondary that is shown in the upper half of Fig. 13-
11 from supplying charge to the batteries. We require that, during normal
operation, the voltage at point A in Fig. 13-11 be greater than the battery
voltage so that the batteries are charged. One way to do this is to specify
the transformer to have two identical secondary coils. The voltage dropped
across the collector-emitter junction of Q_1 (as well as across R_1) will make
the battery voltage less than the voltage at point A.

CRITICAL LOADS
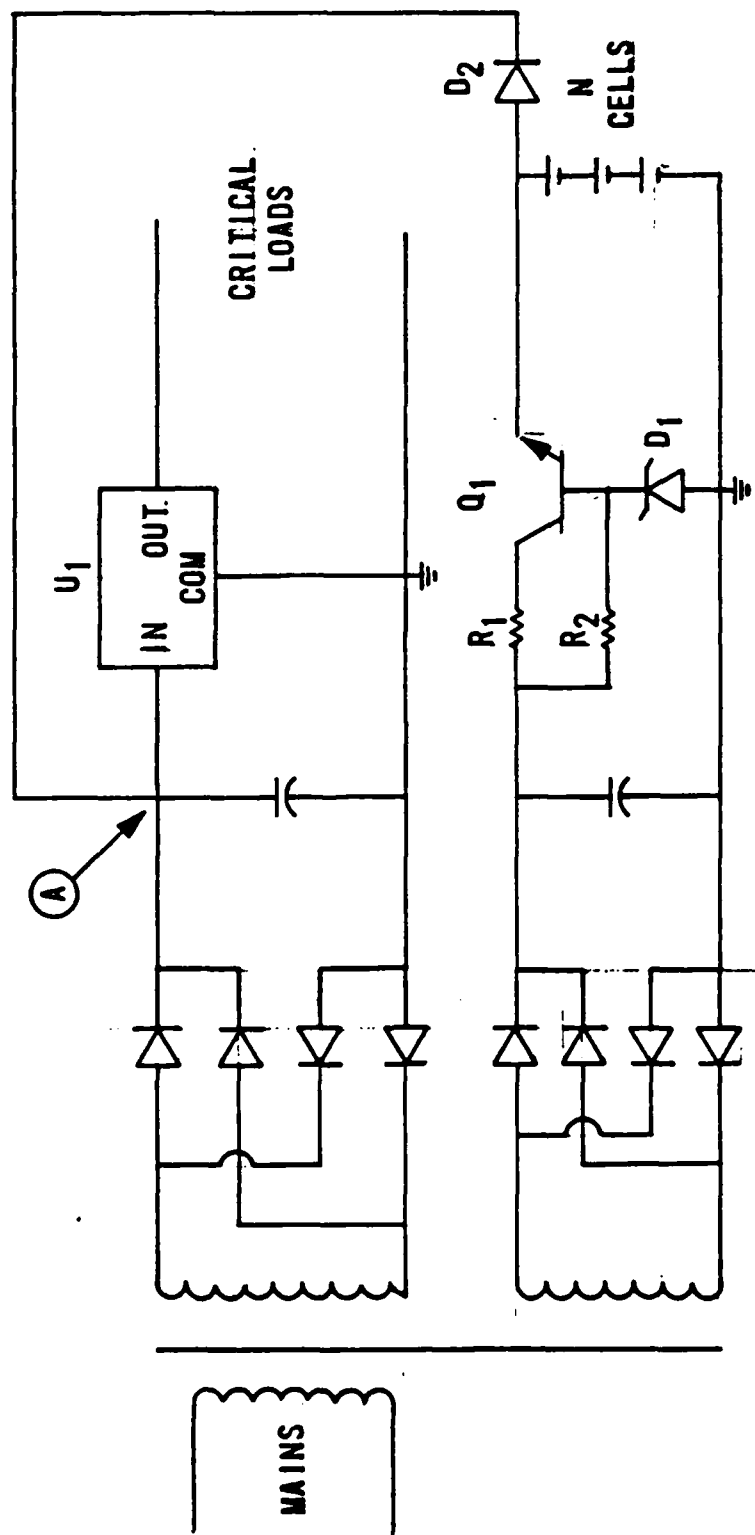
U₁ IN OUT. COM

A

Q₁ R₁ R₂ D₁

D₂ N CELLS

MAINS

Fig.13-11

PROTECTION OF CIRCUITS AND SYSTEMS FROM UPSET
UPSET THRESHOLD

Nearly all transient protection circuits allow a small fraction of the incident transient, called the "remnant," to propagate to the protected devices. In a properly designed protection circuit the remnant will have insufficient energy, current, or voltage to damage protected devices. However, there is still concern that the remnant could be misinterpreted as valid data. Such misinterpretation is called "upset." The threshold for upset is often within the normal range of input voltages to the system. Therefore, one can not discriminate against upset on the basis of voltage levels alone.

UPSET AVOIDANCE

We now discuss some practical suggestions that can be used to reduce the possibility of upset.

Brown, et al. (1973) recognized that decreasing the response of the circuit to high-frequency signals would decrease the vulnerabilty of the circuit to upset by fast transients such as EMP. There are four particularly helpful suggestions that can be made :

1. A low-pass filter (with a cutoff frequency of 10 kHz or less) can be inserted in series with the input data line.

2. Analog interface devices with the smallest acceptable gain-bandwidth product should be used.

3. Relatively slow digital logic devices should be used in interface

circuits.

4.   Edge-triggered logic should be avoided in digital interface circuits.

Brown, et al. (1973) also suggested the use of error-detection codes in digital transmissions to decrease the vulnerabilty to upset. This is a special case of bandwidth reduction, since transmitting redundant information in an error-detection code reduces the amount of information that can be transmitted per unit time.

Use of balanced, differential line and input amplifier with both large common-mode rejection ratio (CMRR) at frequencies above 1 MHz and a large range of permissible common-mode input voltage will decrease vulnerability to common-mode transients. While this is a good suggestion, one must realize that it does nothing to avoid upset from differential-mode transients.

DIGITAL CIRCUITS

Digital circuits appear to be more easily protected against upset than analog circuits. There are two reasons why this is so:

1.   The voltage in a digital circuit has only two valid states, high and low. Values of voltage between these two states can be recognized as inappropriate. This is in contrast to an analog circuit for which a proper signal voltage is allowed to vary continuously over some range. In practice, many analog signals when converted to a digital format require 12 bits (which is equivalent to 4096 different states). If the analog signal has a range of 20 volts (e.g. -10 to +10 volts), the voltage increment per state is about 0.005 volts. This is a much smaller

value per state than for digital signals, where a margin of 1 volt (or more) separates the two states.

2. Digital data can be easily stored in a memory, and transmitted redundantly with a large time delay between transmissions. If both transmissions have identical content, there is a negligible chance that a transient corrupted both transmissions in the same way. There is no convenient analog memory that would permit a corresponding operation.

We can reject noise on digital lines by using interface devices with Schmitt-trigger inputs. A typical 7414 or 74LS14 Schmitt-trigger digital input requires that the input signal have a change in voltage of at least 0.8 volts before the output voltage will change. This allows the system to reject noise that has a magnitude less than this value.

The use of a low-pass filter with digital logic is particularly simple if logic circuits of the complementary metal-oxide semiconductor (CMOS) family are used. The large input impedance of the CMOS gate, which can be modeled as a 10^{12} Ω resistance in parallel with a 5 pF capacitance, constitutes a negligible load for the output of a simple RC low-pass filter as shown in Fig. 14-1. The value of R can be as large as 1 MΩ without introducing complications in the circuit design. If electrolytic capacitors are used for C, solid tantalum units with a capacitance value of less than about 10 μF would be preferred, owing to concerns about DC leakage current. The two power supply connections to the CMOS logic are labeled V_{DD} and V_{SS} in Fig. 14-1, where V_{DD} is the more positive supply. Typical values are

$$V_{DD} = 5 \text{ to } 10 \text{ volts} \quad \text{and} \quad V_{SS} = 0 \text{ (ground)}$$

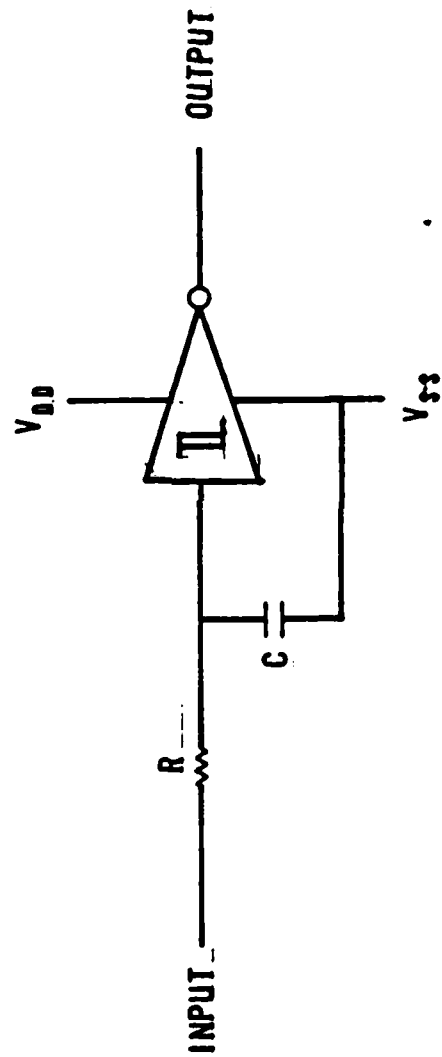The output from the low-pass filter no longer has the sharp edges that are

Fig.14-1

characteristic of a good digital signal. Therefore, the output of the low-pass filter must be connected to the input of a Schmitt-trigger gate to properly interpret the degraded digital signal. Incidentally, a large value of R will provide substantial overvoltage protection for the CMOS input, as discussed in Chapter 11.

Brown, et al. (1973) suggested the use of a delay-line and an AND gate to verify that digital input data are stable. A Schmitt-trigger input circuit must be used for the logic device shown in Fig. 14-2, owing to the degredation in rise and fall times of the signal by the low-pass filter. The Schmitt-trigger input also provides noise immunity as discussed above. We can combine this idea with the use of Schmitt-trigger inputs to obtain the circuit shown in Fig. 14-2. The output of the circuit in Fig. 14-2 is high only if the input has been stable in the high state. This circuit helps avoid upset from transients that cause a high state to become a low state. When this circuit's output is in the low state, one can not tell if the input signal is supposed to be in the low state, or whether a high state has been corrupted by a negtive-going transient. Thus this circuit is appropriate for use in systems where a high state indicates a critical operation (e.g. launch weapons) and the low state indicates a routine or benign condition.

Valid data on a balanced digital line requires that the two lines have complementary states. Buurma (1978) described the use of an exclusive-or gate with Schmitt-trigger inputs to detect non-valid conditions on a balanced digital data line, as shown in Fig. 14-3. Whenever the output of the exclusive-or gate is in the low state, the data are invalid. Conversely, when the output of the exclusive-or gate is in the high state, the data are valid.
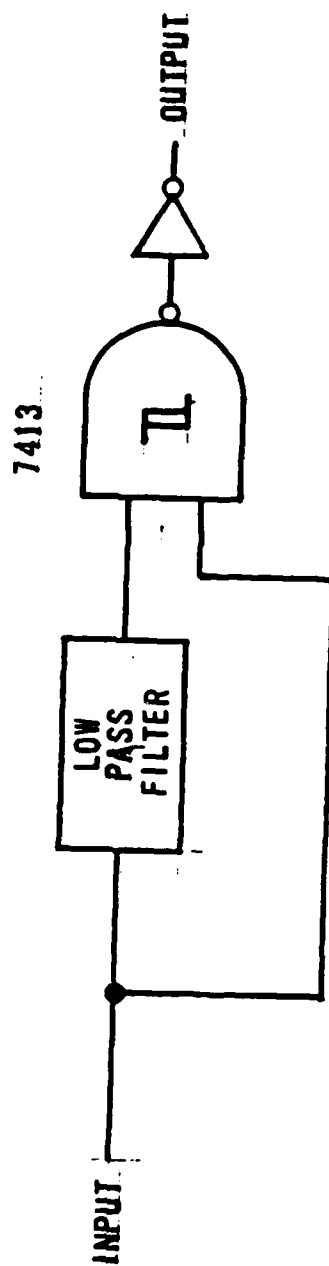
INPUT

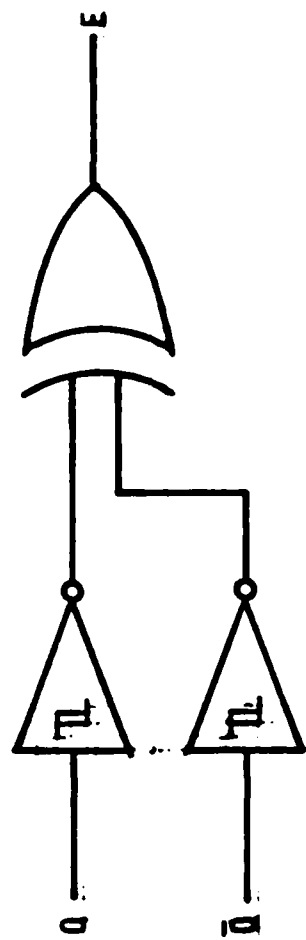LOW
PASS
FILTER

7413

1

OUTPUT

Fig.14-2

Fig. 14-3

The circuit in Fig. 14-3 is not presently available in a single integrated circuit package, but it certainly could be fabricated in a single package if the market were to demand it.

COORDINATION OF TRANSIENT PROTECTION AND UPSET AVOIDANCE

Since transient protection is required on all interface circuits, it is reasonable to consider including a transient detection circuit with the protection circuit(s). The output of the overvoltage detection circuit would be used to inhibit acceptance of data during or immediately after a transient.

Knight (1972) suggested a novel transient protection circuit similar to that shown in Fig. 14-4. The avalanche diode and light emitting diode (LED) serve as a transient detection circuit. The light from the LED causes the switch on the right hand side of Fig. 14-4 to close, thus protecting the load. The delay line compensates for the response time of the photodetector and switch. While Knight (1972) advocated this circuit for protection applications, it is a simple modification to remove the shunt switch and use the output of the photodetector to indicate the presence of overvoltages. This output could be used to inhibit the acceptance of data.

The circuit shown in Fig. 14-4 responds only to positive overvoltages, however bipolar circuits are a simple extension, as shown in Fig. 14-5. In the circuit of Fig. 14-5, avalanche diodes D_1 and D_2 provide the final stage in a transient protection circuit (for simplicity, the earlier stage(s), which may include a spark gap) are not shown. Diodes D_1 and D_2 are essential, because we can not rely on diodes D_3, D_4 for protection. The shunt path

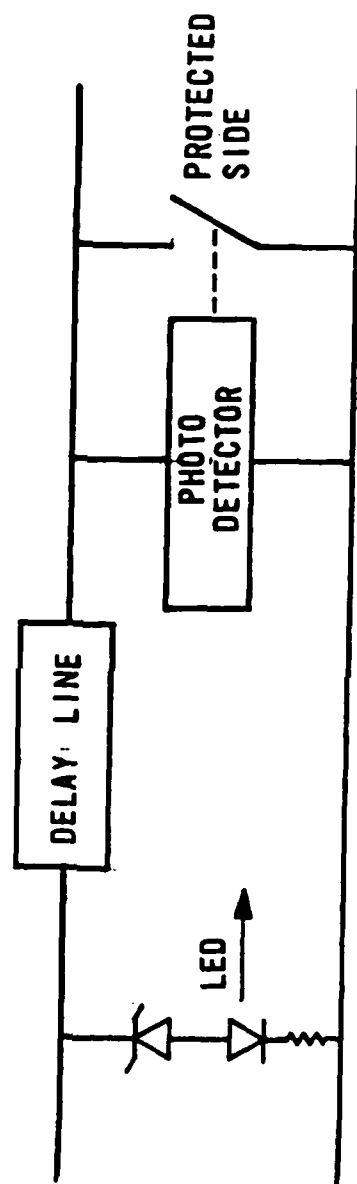# OPTICAL TRANSMISSION PATH CIRCUMVENTS FINITE RESPONSE TIME OF SWITCH (KNIGHT, 1972)
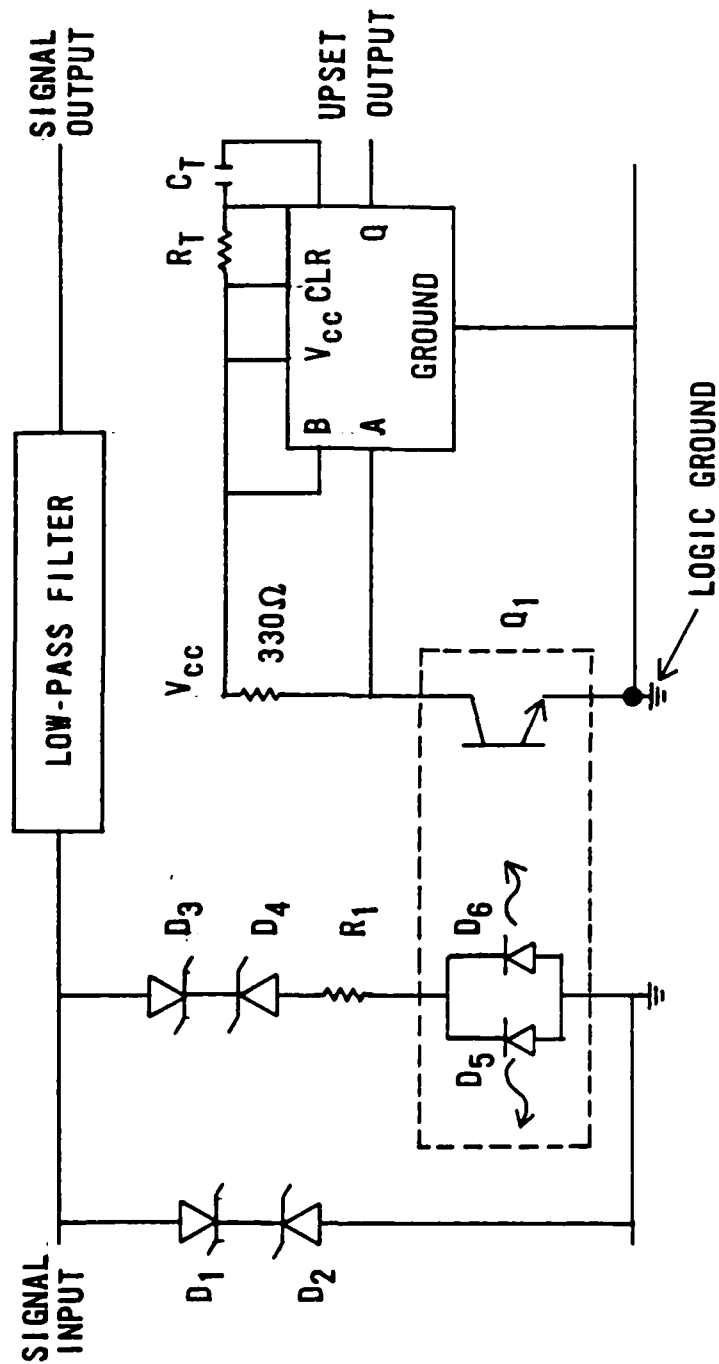

Fig. 14-4

Fig. 14-5

through D_3 and D_4 includes a resistance R_1, and probably considerable parasitic inductance, that defeats any protective function. The pair of LEDs, D_5 and D_6, along with phototransistor Q_1, are contained inside a single package, e.g. model H11AA1 (which is available from General Electric, Motorola, or Siemens/Litronix) or model OPI2500 (which is available from TRW Optron).

The value of R_1 is determined so that the LEDs are not damaged by transient pulses. The circuit will be designed to operate the LEDs, D_5 and D_6, at 10 mA and 1.1 V. We let V_U denote the magnitude of the voltage at which the upset detection circuit will respond. Of course, the magnitude of the clamping voltage of the series combination of protection diodes D_1 and D_2 must be greater than V_U, else D_3 and D_4 will not conduct. We let V_Z denote the breakdown voltage of the series combination of diodes D_3 and D_4. From Kirchhoff's voltage law, we obtain:

$$V_Z + (10 \text{ mA} \times R_1) + 1.1 \text{ volts} = V_U$$

This equation has two unknowns, V_Z and R_1.

We need to be certain that the LEDs will not be damaged by transient currents. We let V_P denote the magnitude of the clamping voltage of the series combination of protection diodes D_1 and D_2. The maximum current in either LED should not exceed about 60 mA, this occurs at about 1.3 volts across the LED. We can apply Kirchhoff's voltage law again and obtain:

$$V_Z + (60 \text{ mA} \times R_1) + 1.3 \text{ volts} = V_P$$

This equation also has two unknowns, V_Z and R_1. We can solve these two equations simultaneously to complete the design.

When one of the LEDs is illuminated, the phototransistor conducts, and the voltage at the A input terminal of the 74221 monostable multivibrator goes to the low state. This triggers the multivibrator, which provides a positive pulse at its output terminal. The pulse duration is a function of a timing resistance, R_T, and capacitance, C_T. The multivibrator automatically resets itself at the end of the pulse.

Unfortunately, optoisolators with a phototransistor output have a slow response. The low-pass filter or delay line in the circuit shown in Fig. 14-5 should have a delay time of several microseconds in order to be certain that the output of the upset detection circuit has responded before the remnant has propagated to the output port of the delay line.

One could substitute an *optoisolator* with a photodiode output, and decrease the response time to less than 0.1 μs. While, this greatly increases the cost of the optoisolator, the total circuit cost may decrease owing to the smaller delay line that can be used with a faster optoisolator. Suitable optoisolators with a pair of photodiodes and internal amplifiers include the Hewlett-Packard HCPL-2630.

An alternate transient detection circuit might have a PIN photodiode, D_1, detect luminosity from a spark gap, as shown in Fig. 14-6. The signal from the photodiode is detected with a high-speed comparator. Resistors R_2 and R_3 bias the comparator so that the output is normally in the low state. When the photodiode conducts, the voltage across R_1 is greater than the voltage across R_3, and the output of the comparator switches to the high state. Typical values of the resistors are
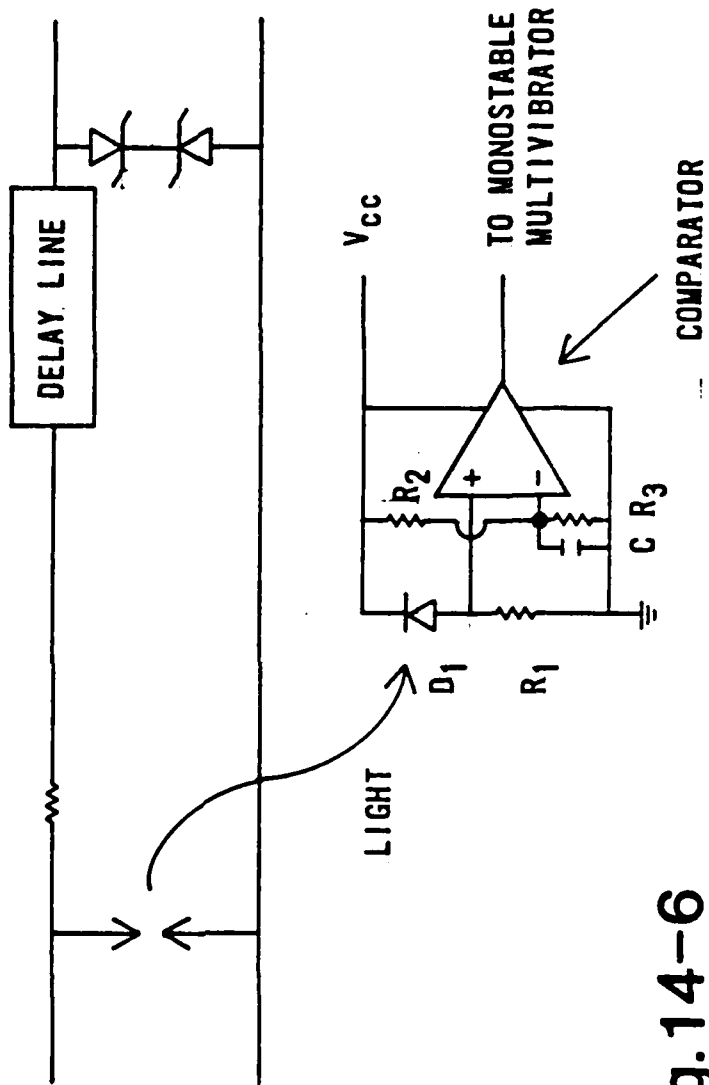
Fig. 14-6

R_1 = 2 kΩ R_2 = 4.7 kΩ R_3 = 1 kΩ

The capacitor C is a 0.01 μF ceramic bypass capacitor that is included to insure a stable voltage at the inverting input of the comparator.

The spark gap in the circuit in Fig. 14-6 must have a transparent glass case (e.g. Siemens button type gap) in order to flood the photodiode with intense light during a transient. Also, the photodiode should be specified to have adequate response to blue light, since this color predominates when the spark gap operates in the arc region.

TRANSIENT DETECTION FOR MAINS

Transients on the mains can be detected with the circuit shown in Fig. 14-7. The varistor, V_1, is a metal oxide varistor with a clamping voltage of less than 1 kV at the maximum expected surge current. Avalanche diodes D_1 and D_2 may have a nominal breakdown voltage of 200 volts and a steady-state power rating of at least 5 watts. These diodes provide the discrimination between normal mains voltages and a transient overvoltage. Avalanche diodes D_3 and D_4, and resistor R_2, protect the LEDs in the optoisolator from excessive current. Suitable types of avalanche diodes for D_3 and D_4 have a breakdown voltage of 6.8 volts and a steady-state power rating of 0.5 watt. Diodes D_3 and D_4 do not need to have large power ratings, since the current in them is less than the current in diodes D_1 and D_2 (owing to the shunt path through the LEDs) and the voltage across D_3 and D_4 is much less than the voltage across D_1 and D_2. The optoisolator is the same type as discussed above in Fig. 14-5, such as model H11AA1.

The value of resistors R_1 is determined by the maximum voltage across the varistor and the maximum tolerable current in D_1 and D_2. Because this circuit is isolated from ground, and the energy deposited in the resistance is large, the resistance has been divided into two units of equal value. If the varistor maintains the voltage at less than 1 kV and D_1 and D_2 can tolerate a current of 25 mA (5 watts/200 volts = 25 mA), then we calculate

$$1 \text{ kV} / 25 \text{ mA} = 2 R_1 \qquad \text{or} \quad R_1 = 20 \text{ k}\Omega$$

These resistors should be either 2 watt carbon composition units or wirewound units with a steady-state power rating of at least 2 watts.

Resistor R_2, along with diodes D_3 and D_4, protects the LEDs from excessive current. The maximum allowable steady-state LED current, about 50 mA, is greater than the maximum steady-state current in diodes D_1 and D_2. Therefore, we can let all of the current in D_1 and D_2 pass through the LEDs. This maximizes the current in the LEDs, and thus makes the optoisolator faster responding. The series combination of diodes D_3 and D_4 will not conduct at a voltage of less than $((6.8 \times 0.95) + 0.6)$ volts, or about 7.0 volts. (We take into account a $\pm$ 5% tolerance on the 6.8 volt breakdown potential.) We determine the value of R_2 so that when 25 mA passes through D_1 and D_2, there is just 7.0 volts across D_3 and D_4. Since the LED has a drop of about 1.1 volts when conducting, we obtain $R_2 = 240 \ \Omega$.

The output of the phototransistor in Fig. 14-7 can be connected to a multivibrator as was shown in Fig. 14-5. The presence of low-pass filters downstream from the varistor is standard in mains transient protection circuits (see Chapter 12). These low-pass filters will delay the remnant. Tests need to be performed in the laboratory to verify that the delay time of
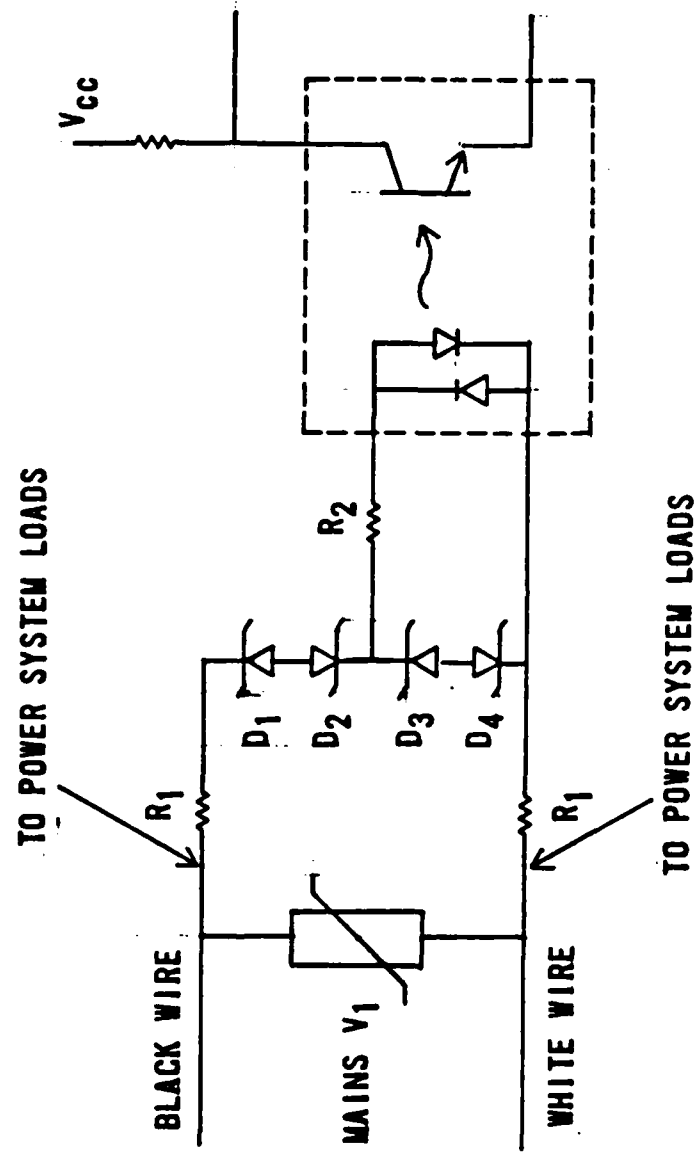
Fig.14-7

the low-pass filter is at least as long as the response time of the transient detection circuit of Fig. 14-7.

APPLICATION OF UPSET AVOIDANCE CIRCUITS

In the circuits shown in Figs. 5 through 7, a delay line or low-pass filter is necessary between the transient sensor and the connection of the input line to the system, in order that the system is inhibited before the peak of the transient reaches the system. In addition to inhibiting the system, one might also activate an electro-mechanical latching mechanism that would signal maintenance crews to check for possible damage caused by the transient.

Since the transient sensors in Figs. 5 through 7 depend upon voltage discrimination, there will be a "window of susceptibility" for magnitudes of input voltage that exceed the upset threshold but are less than the level needed to activate the transient sensor. This window of susceptibility makes the upset problem very difficult to solve.

It is important to recognize that not all input conductors to a system need to be connected to a transient sensor. Input lines should be connected to a transient sensor if they are either:

1. connected to transducers that are illuminated by exterior electromagnetic fields or exposed to direct injection of current by lightning,

2. carry data of a particularly critical nature, or

3. have relatively lengthy or poorly shielded transmission lines.

The outputs of multiple transient sensors should be connected to an OR gate so that we obtain a single digital signal to inhibit the system.

Hardening a circuit to prevent upset needs to be done during the design of the equipment to be protected. Effective upset hardening is unlikely to be accomplished by later connecting a box to each input. There is general agreement that hardening a system against upset is much more difficult than protecting it from destructive transients.

UPSET DUE TO INTERRUPTION OF MAINS

A system can also be upset by transients on power lines. If spark gaps are used for transient overvoltage protection of power lines, the line current must be interrupted for several cycles to stop the "power-follow" in the spark gap (see discussion of spark gaps). This is why the lights in buildings flicker or blink during thunderstorms: lightning hits an overhead power line, drives spark gap(s) into the arc mode (which puts an effective short-circuit across the line), and the arc is interrupted by automatically reclosing circuit breakers.

To avoid these brief losses of AC power, one could prohibit the use of spark gaps (and other devices which inherently have a "follow-current" problem) as SPDs on power lines. Metal oxide varistors offer an effective substitute for small and moderate surges. If the power line is buried underground, the very large surges that are associated with direct lightning strikes are largely avoided. However, for overhead power lines, spark gaps

appear to be necessary for economical protection of the distribution system. Moreover, there are natural spark gaps such as insulator flash-over to ground and arcs between the line and adjacent trees.

Digital computers are particularly susceptible to upset from unanticipated interruption of power. During the fractions of a second that power is interrupted by the automatic circuit breakers, the filter capacitors in the DC power supplies inside the computer will be the only source of power to keep the computer's memory opertional. Significant upset protection can be obtained by making these filter capacitors large enough to operate the entire computer for at least one second. Alternately, an "uninterruptible power supply" (UPS) can be connected between the power line and the computer. The operation of an UPS is described in Chapters 12 and 13 for AC and DC circuits, respectively. A typical commercial grade UPS that can supply 3 A at 120 Vrms for 20 minutes costs about $1500 in small quantities. While this is more expensive than increasing the size of the filter capacitors in the DC supply, it does offer protection against sustained power interruptions.

A third technique to avoid upset due to temporary loss of AC power is to use a voltage-responsive upset detection circuit to switch the load into a stand-by state. For example, computers with CMOS static random access memory (RAM) can retain their data for weeks or months when powered from small Ni-Cd or Li batteries that can be mounted on the printed circuit board that contains the memories. This technique is used in the "continuous memory" feature of some handheld calculators and is available as an option on a few mainframe computers.

REFERENCES

Andrich, E., "Properties and Applications of PTC Thermistors," _Electronics Applications Bulletin_ (Netherlands), 26:123-144, 1966.

Bazarian, A., "Gas Tube Surge Arresters for Control of Transient Voltages," _Rome Air Development Center 1980 Electrostatic Discharge Symposium_, pp.44-53, 1980.

Bell Laboratories, _EMP Engineering and Design Principles_, Whippany, NJ, 151 pp., 1975.

Bennison, Eric, A. J. Ghazi, and P. Ferland, "Lightning Surges in Open Wire, Coaxial, and Paired Cable, _IEEE Trans. Communications_, COM-21:1136-1143, October 1973.

Bodle, David W., and James B. Hayes, "Lightning Protection Circuit," U.S. Patent 2,789,254, 16 April 1957.

Bodle, David W., and P. A. Gresh, "Lightning Surges in Paired Telephone Cable Facilities," _Bell System Tech. J._, 40:547-576, March 1961.

Bowling, Edward L., "Electrolytic Capacitors," U.S. Patent 4,141,070, 20 February 1979.

Brainard, John P., and Larry A. Andrews, "Dielectric Stimulated Arcs in

Lightning-Arrestor Connectors," IEEE Trans. Components, Hybrids, and Mfg. Technology, CHMT-2:309-316, September 1979.

Brainard, John P., Larry A. Andrews, and Robert A. Anderson, "Varistor-Initiated Arcs in Lightning Arrestor Connectors," IEEE 1981 Conference on Electronic Components, Atlanta, Georgia, pp.308-312, May 1981.

Brook, M., N. Kitagawa, and E. J. Workman, "Quantitative Study of Strokes and Continuing Currents in Lighting Discharges to Ground," J. Geophysical Res., 67:649-659, 1962.

Brown, R. M. and 7 others, "EMP Electronic Design Handbook," U.S. Air Force Weapons Laboratory AFWL-TR-74-58, DDC-AD-918277, April 1973.

Burger, J.R., "Protection Against Junction Burnout by Current Limiting," IEEE Trans. Nuclear Science, NS-21:28-30, October 1974.

Buurma, Gerald, "CMOS Schmitt Trigger--a Uniquely Versatile Component," National Semiconductor Application Note 140, reprinted in National Semiconductor CMOS Databook, 1978

Campi, Morris, "EMP Line Filter using MOV Devices," U.S. Patent 4,021,759, 3 May 1977.

Cergel, L., "General CMOS Characteristics," chapter in Motorola M^C Mos Handbook, 1974.

Clark, O. Melville, and R. D. Winters, "Feasibility Study for EMP Terminal Protection," General Semiconductor Industries Report TPD003 for U.S. Army Material Command, DDC-AD-909267, March 1973.

Clark, O. Melville, "Suppression of Fast Rise-time Transients," IEEE EMC Conference, pp.66-71, 1975.

Clark, Oscar M., "Power Surge Protection System," U.S. Patent 3,934,175, 20 January 1976.

Clark, Oscar Melville, "Four Terminal Pulse Suppressor," U.S. Patent 4,325,097, 13 April 1982.

Chowdhuri, P., "Breakdown of P-N Junctions by Transient Voltages," Direct Current, 10:131-139, August 1965.

Chowdhuri, P., "Circuit for Protecting Semiconductors Against Transient Voltages," U.S. Patent 3,793,535, 19 February 1974.

Cohen, E. J., J. B. Eppes, and E. L. Fisher, "Gas Tube Arresters," IEEE 1972 International Communications Conference Proceedings, paper 43, 1972.

Cooper, Howard K., "Forward Transient Response of Silicon Diffused P-N Junctions," Proc. National Electronics Conference, 18:107-113, 1962.

Cushman, Lester A., "Lightning Arrester," U.S. Patent 2,922,913, 26 January 1960.

Damljanovic, Dragoljub, and Vojislav Arandjelovic, "Input Protection of Low Current DC Amplifiers by GaAsP Diodes," J. Phys. E (Scientific Instruments), 14:414-417, 1981.

de Souza, Alwyn A., "Surge Protection Circuit," U.S. Patent 3,353,066, 14 November 1967.

Doljack, Frank A., "Polyswitch PTC Devices," IEEE Trans. Components, Hybrids, and Mfg. Technology, CHMT-4:372-378, December 1981.

Domingos, Henry, and Donald C. Wunsch, "High Pulse Power Failure of Discrete Resistors," IEE Trans. Parts, Hybrids, and Packaging, PHP-11:225-229, September 1975.

Enlow, Ed W., "Determining an Emitter-Base Failure Threshold Distribution of NPN Transistors," Rome Air Development Center 1981 Electrostatic Discharge Symposium, p.145-150, 1981.

Erickson, John, "Lightning and High Voltage Surge Protection for Balanced Digital Transmission Devices," U.S. Army Electronics Command ECOM-4027, NTIS-AD-752448, September 1972.

Fisher, F. A., "Overshoot: A Lead Effect in Varistor Characteristics," General Electric Company, Schenectady, NY, Corporate Research and Development Report 78CRD201, September 1978.

Fuquay, D.M., A.R. Taylor, R.G. Hawe, C.W. Schmid, "Lighting Discharges that Caused Forest Fires," J. Geophys. Res., 77:2156-8, April 1972.

Glasstone, Samuel, and Philip J. Dolan, The Effects of Nuclear Weapons, Third Edition, U.S. Government Printing Office: Washington DC, 653 pp., 1977.

Greenwood, Allan, Electrical Transients in Power Systems, New York: Wiley, 544 pp., 1971.

Grossner, Nathan R., Transformers for Electronic Circuits, 2nd edition, New York: McGraw-Hill, 467 pp., 1983.

Gutzwiller, Frank W., "Protective Control Circuits," U.S. Patent 3,213,349, 19 October 1965.

Harnden, J. D., F. D. Martzlof, W. G. Morris, and F. G. Golden, "Metal-Oxide Varistor: A New Way to Suppress Transients," Electronics, p91-95, 9 October 1972.

Institute of Electical and Electronics Engineers, IEEE Guide for Surge Voltages in Low-Voltage AC Power Circuits, IEEE Standard 587-1980, ANSI Standard C62.41-1980, 1980.

Kawiecki, Chester J., "Surge Protector," U.S. Patent 3,564,473, 16 February 1971.

Kawiecki, Chester J., "Spark Gap Device having a Thin Conductive Layer for Stablizing Operation," U.S. Patent 3,588,576, 28 June 1971.

Kawiecki, Chester J., "Spark Gap Device," U.S. Patent 3,811,064, 14 May 1974.

Knight, Stephen, "Circuit Protection Apparatus Utilizing Optical Transmission Path," U.S. Patent 3,648,110, 7 March 1972.

Knox, K.A.T., "Semiconductor Devices in Hostile Electrical Environments," Electronics and Power, 19:557-560, December 1973.

Kreider, E.P., C.D. Weidman, and R.C. Noggle, "The Electric Field Produced by Lightning Stepped Leaders," Journal of Geophysical Research, 82:951-960, 20 Feb 1977.

Lasitter, H.A., and D.B. Clark, "Nuclear Electromagnetic Pulse Protective Measures Applied to a Typical Communications Shelter," Naval Civil Engineering Laboratory TN-1091, DDC AD-707-696, April 1970.

Lennox, C. R., "Experimental Results of Testing Resistors Under Pulse Conditions," Sandia Laboratories SC-TM-67-559, November 1967.

Lesinski, Leon C., "Electronic Surge Arrestor," U.S. Patent 4,390,919, 28 June 1983.

Longmire, Conrad L., "On the Electromagnetic Pulse Produced by Nuclear Explosions," IEEE Trans. Electromagnetic Compatibility, EMC-20:3-13, Feb 1978.

Lopez, C., A. Garcia, E. Munoz, "Deep-level Changes Associated with the Degradation of Ga $As_{0.6}$ $P_{0.4}$ LEDs," Electronic Letters, 13:459-461, 4 August 1977.

M^C Neill, Ralph, "Lightning Arrester," U.S. Patent 1,382,795, 28 June 1921.

Martzloff, Francois D., and Gerald J. Hahn, "Surge Voltages in Residential and Industrial Power Circuits," IEE Trans. Power Apparatus and Systems, PAS-89:1049-1056, July 1970.

Martzloff, F. D., "Coordination of Surge Protectors in Low-voltage AC Power Circuits," IEEE Trans. Power Apparatus and Systems, PAS-99:129-133, January 1980.

Martzloff, F.D., "The Coordination of Transient Protection for Solid-State Power Conversion Equipment," 1982 IEEE/Industry Applications Society International Semiconductor Power Conference, 24 May 1982. (Reprinted as General Electric Corporate Research and Development publication 82CRD171, Schenectady, NY, June 1982)

Martzloff, F. D., "The Propagation and Attenuation of Surge Voltages and Surge Currents in Low-voltage AC Circuits," IEEE Transactions on Power Apparatus and Systems, PAS-102:1163-1170, May 1983.

Martzloff, F.D., "Matching Surge Protective Devices to Their Environment," IEEE/Industry Applications Society meeting, October 1983. (Reprinted as General Electric Corporate Research and Development publication 83CRD169, Schenectady, NY, July 1983)

Matsuoka, Michio, Takeshi Masuyama, and Yoshio Iida, "Non-linear Resistors," U.S. Patent 3,496,512, 17 February 1970.

Morrison, Ralph, Grounding and Shielding Techniques in Instrumentation, Second Edition, John Wiley and Sons: New York City, 146 pp., 1977.

Ott, Henry W., Noise Reduction Techniques in Electronic Systems, John Wiley and Sons: New York City, 294 pp., 1976.

Palmer, James R., "Power and Lightning Surges in Coaxial Distribution Systems," IEEE 1977 Cable Television Reliability Symposium, p.27-31, 1977.

Person, Herman R., "Electrical Surge Arrestor," U.S. Patent 3,480,832, 25 November 1969.

Philip, Herbert R., Lionel M. Levinson, "ZnO Varistors for Protection Against Nuclear Electromagnetic Pulses," J. Appl. Phys., 52:1083-90, February 1981.

Poindexter, Carlson H., "Electrical Protective Device using a Reed Relay," U.S. Patent 3,858,089, 31 December 1974.

Popp, E., "Lightning Protection of Line Repeaters," Entwicklungs Berichte der

Siemens Halske Werke, 31:25-28, September 1968.

Pujol, H.L., "COS/MOS Electrostatic Discharge Protection Networks," RCA
Application Note ICAN-6572, reprinted in RCA COS/MOS Integrated Circuit Data
Book, 1977.

Reynolds, J.E., "Low Cost Transient Suppressors: Exploratory Development,"
U.S. Army Electronics Command ECOM-0325-F, DDC-AD-743996, May 1972.

Richman, Peter, "Single Output, Voltage and Current Surge Generation for
Testing Electronic Systems," IEEE 1983 Electromagnetic Compatibility
Symposium, p.47-51, 1983.

Ricketts, L.W., J.E. Bridges, J. Milletta, EMP Radiation and Protective
Techniques, New York:Wiley, 380pp., 1976.

Rovere, Lewis H., Phillip H. Estes, and Joseph W. Milnor, "Electrical
Protective Device, U.S. Patent 1,971,146, 21 August 1934.

Saburi, O., and K. Wakino, "Processing Techniques and Applications of Positive
Temperature Coefficient Thermistors," IEEE Trans. Component Parts, CP-10:53-
67, June 1963.

Sandia Laboratories, "Electromagnetic Pulse Handbook for Missiles and Aircraft
in Flight," Sandia Labs SC-M-71-0346, September 1972.

Sherwood, Robert A., "Protecting CATV Transmission Equipment from Surges,"

IEEE 1977 Cable Television Reliability Symposium, pp.35-38, 1977.

Singletary, John B., and John A. Hasdal, "Methods, Devices, and Circuits for the EMP Hardening of Army Electronics," U.S. Army Electronics Command ECOM-0085-F, DDC-AD-885224, June 1971.

Smith, M.N., "Practical Application and Effectiveness of Commercially Available Pulse Voltage Transient Suppressors," U.S. Naval Civil Engineering Laboratory TN-1312, DDC AD-773-074, December 1973.

Smith, Marvin W., and M. D. M^CCormick, Transient Voltage Suppression Manual, Third Edition, General Electric Company, Auburn, NY, 162 pp., 1982.

Smithson, A. K., "Lightning Protection for W.R.E. Computer Cable Network," Weapons Research Establishment, Adelaide, South Australia, WRE-TM-1796, DDC-AD-A047080, May 1977.

Stansberry, C. L., "EMP, Lightning, and Power Transients: Their Threat and Relationship to Future EMP Standards," National Communications Systems TN-22-77, DDC-AD-A058404, November 1977.

Tasca, Dante M., "Pulse Power Response and Damage Characterization of Capacitors," Rome Air Development Center 1981 Electrostatic Discharge Symposium, p.174-191, 1981.

Thomas, M.E., F.L.Pitts, "Direct Strike Lightning Data," NASA-TM-84-626, March 1983.

Tucker, T. J., "Spark Initiation Requirements of a Secondary Explosive," Annals of the New York Academy of Science, 152:643-653, October 1968.

Trybus, P. R., A. M. Chodorow, D. L. Endsley, and J. E. Bridge, "Spark Gap Breakdown at EMP Threat Level Rates of Rise," IEEE Trans. Nuclear Science, NS-26:4959-4963, December 1979.

Uman, Martin, Lightning, New York: M^CGraw-Hill, 264pp., 1969.

Uman, M., and E.P. Kreider, "A Review of Natural Lightning," IEEE Transactions on Electromagnetic Compatibility, EMC-24:79-112, May 1982.

United States of America, Department of Defense, Civil Preparedness Agency TR-61A, "EMP Protection for Emergency Operating Centers," July 1972. (Reprinted as Lawrence Livermore Laboratory Report PEM-8)

United States of America, Department of Defense Military Standard 188-124, "Grounding, Bonding, and Shielding for Common Long Haul/Tactical Communication Systems," 14 June 1978.

United States of America, Department of Defense Military Handbook 419, "Grounding, Bonding, and Shielding for Electronic Equipments and Facilities," 21 January 1982.

Van Keuren, E., "Effects of EMP Induced Transients on Integrated Circuits," IEEE 1975 Electromagnetic Compatibility Symposium, paper 3AIIe, 1975.

Vance, E.F., "EMP-induced Transients in Long Cables," *IEEE* 1975 *Electromagnetic Compatiibility Symposium*, paper 3AIIa, 1975.

Voorhoeve, Ernst W., "Power-Limiting Electrical Barrier Device," *U.S.Patent* 3,878,434, 15 April 1975.

Williams, Donald P., Marx Brook, "Magnetic Measurements of Thunderstorm Currents," *J. Geophys. Res.*, 68:3243-7, May 1963.

Williams, R. L., "Test Proceedures for Evaluating Terminal Protection Devices used in EMP Applications," U.S. Army Harry Diamond Laboratories HDL-TR-1709, DDC-AD-A019098, June 1975.

Woody, J.A., "Modeling of Parasitic Effects in Discrete Passive Components," Rome Air Development Center, RADC-TR-83-32, February 1983.

Wunsch, D.C., and R. R. Bell, "Determination of Threshold Failure Levels of Semiconductor Diodes and Transistors Due to Pulse Voltages," *IEEE Trans. Nuclear Science*, NS-15:244-259, December 1968.

END

FILMED

12-85

DTIC