NASA-CR-173565
19840017821

The Telecommunications and Data Acquisition Progress Report 42-77

January–March 1984

E.C. Posner
Editor

May 15, 1984

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

```
D 84N25889
          DISPLAY 84N25889/2
       84N25889*#   ISSUE 16   PAGE 2478   CATEGORY 32   RPT#: NASA-CR-173565 NAS
       1.26:173565 TDA-PR-42-77   84/03/15   164 PAGES   UNCLASSIFIED DOCUMENT
UTTL:  The telecommunications and data acquisition report   TLSP: Progress Report
AUTH:  A/POSNER, E. C.   PAT: A/ed.
CORP:  Jet Propulsion Lab., California Inst. of Tech., Pasadena.     AVAIL.NTIS
       SAP: HC A08/MF A01
MAJS:  /*DATA ACQUISITION/*DEEP SPACE NETWORK/*GEODYNAMICS/*PROJECT SETI/*RADIO
       NAVIGATION/*TELECOMMUNICATION
MINS:  / GROUND STATIONS/ INFORMATION SYSTEMS/ RADIO ASTRONOMY/ VERY LONG BASE
       INTERFEROMETRY
ANN:   Activities in space communication, radio navigation, radio science, and
       ground-based astronomy are reported. Advanced systems for the Deep Space
       Network and its Ground-Communications Facility are discussed including
       station control and system technology. Network sustaining as well as data
       and information systems are covered. Studies of geodynamics,
       investigations of the microwave spectrum, and the search for
       extraterrestrial intelligence are reported. For individual titles, see
       N84-25890 through N84-25905.
```

The Telecommunications and Data Acquisition Progress Report 42-77

January-March 1984

E.C. Posner
Editor

May 15, 1984

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

N84-25889#

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Tracking and Data Systems (OSTDS).

In geodynamics, the publication reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the search for extraterrestrial intelligence (SETI), it reports on implementation and operations for searching the microwave spectrum. The latter two programs are performed for NASA's Office of Space Science and Applications (OSSA).

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

This and each succeeding issue of the TDA Progress Report will present material in some, but not necessarily all, of the following categories:

OSTDS Tasks:

- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft-Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for New Projects
  - Networks Consolidation Program
  - New Initiatives
  - Network Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- GCF Implementation and Operations
- Data and Information Systems

OSSA Tasks:

- Search for Extraterrestrial Intelligence
- Geodynamics
  - Geodetic Instrument Development
  - Geodynamic Science

Discretionary Funded Tasks

Contents

OSTDS TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT–GROUND

STATION CONTROL AND SYSTEM TECHNOLOGY

DSN Systems Implementation
CAPABILITIES FOR NEW PROJECTS

NETWORK SUSTAINING

DATA AND INFORMATION SYSTEMS

OSSA TASKS
Search for Extraterrestrial Intelligence

Arcsecond Positions for Milliarcsecond VLBI Nuclei of Extragalactic Radio Sources, Part II: 207 Sources

D. D. Morabito, R. A. Preston, M. A. Slade
Tracking Systems and Applications Section

D. L. Jauncey
CSIRO, Sydney, Australia

G. D. Nicolson
CSIR, Johannesburg, South Africa

VLBI measurements of time delay and fringe frequency at 2.29 GHz on baselines of 10^4 *km between Deep Space Network stations have been used to determine the positions of the milliarcsecond nuclei in 207 extragalactic radio sources. Estimated accuracies generally range from* $\sim 0''.1$ *to* $\sim 1''.0$*, in both right ascension and declination, with all sources having uncertainties* $< 4''$ *in both coordinates. The observed sources are part of an all-sky VLBI catalog of milliarcsecond radio sources. Arcsecond positions have now been determined for 752 of these sources. Arcsecond positions serve as a useful starting point in the construction of high-precision VLBI reference frames and are also important for unambiguous determination of optical counterparts to compact radio sources.*

I. Introduction

A survey is underway to develop an all-sky catalog of radio sources with milliarcsecond components at 2.29 GHz. This is being accomplished by searching for compact components in known extragalactic sources with intercontinental VLBI baselines (Refs. 1 and 2). We have used the VLBI measurements of time delay and fringe frequency at 2.29 GHz to determine the positions of the milliarcsecond nuclei. Previously we determined the positions of 546 nuclei (Ref. 3). In this article we determine the positions of an additional 206 nuclei and include an improved position for one source (M 104) from Ref. 3. Estimated accuracies in this article range from $0''.1$ to $3''.8$ in both right ascension and declination. The addition of the Australia to South Africa baseline has allowed the survey to continue south of -45° declination. In total we have determined arcsecond positions for 84% of the detected sources.

Arcsecond positions serve as a useful starting point in the construction of a high precision VLBI reference frame as well as allow unambiguous optical identifications. The positions from this article and Ref. 3 are presently being used as a first step in the formation of a precision reference frame of about 200 sources in which relative radio positions should be determined to milliarcsecond accuracy (Ref. 4).

II. The Observations

The observations were performed with pairs of antennas on either California-Australia, California-Spain, or Australia-South Africa baselines (see Table 1) in 16 separate observing sessions between 1978 and 1982. A list of experiments appears in Table 2. The observations were performed at 2.29 GHz with MK II VLBI recording systems as described in Ref. 3. The length of each observation was a few minutes with 38% of the sources being observed more than once.

III. Method of Position Determination

The details of position determination for this set of sources closely follow the analysis given in Ref. 3, except for three differences:

(1) The data for the 207 sources reported here were reduced using the Caltech/JPL VLBI Mark II correlator and post-correlation software, whereas the data in Ref. 3 were reduced using the NRAO Mark II correlator and post-correlation software.

(2) Increased accuracy in modeling of the troposphere and earth rotation parameters yielded a smaller scatter in the time delay and fringe frequency residual. Typical residual values for the rms time delay scatter were ∼10 ns and for the rms fringe frequency scatter ∼0.1 mHz. The observed scatter was assumed to be an estimate of the random measurement error. Earth rotation corrections were obtained from final smoothed values of UT1-UTC and polar motion from Bureau International de l'Heure (BIH).

(3) Due to improved modeling of the troposphere and earth rotation parameters, 0″.3 rather than 0″.5 was added in a root-sum-square manner to formal estimates of right ascension and declination uncertainty in each experiment to account for unmodeled effects.

As in Ref. 3, observations of sources with well-known positions allowed instrumental delays and frequency offsets to be determined. At least 3 such calibration sources per experiment were spread in time among the sources whose positions we wished to determine (see Table 2). A list of the 44 calibration sources utilized appears in Table 3. Twenty-eight of the calibration source positions have been determined with VLBI (Ref. 5) and can be referred to the FK4 reference frame (Ref. 6) with an accuracy of 0″.1. Other calibration source positions came from 1) Ref. 7, 8 sources, 0″.1 accuracy; 2) Ref. 3, 5 sources, 0″.5 accuracy, 3) Ref. 8, 2 sources, 1″ accuracy, and 4) D. L. Jauncey et al. (private communication), 1 source, 0″.1 accuracy.

IV. Results

The calculated positions of 207 sources and the corresponding uncertainties are shown in Fig. 1. The positions are referred to the equinox of 1950.0, and elliptical aberration terms are included so as to agree with past astronomical convention. The source positions and position uncertainties for sources which were multiply observed were estimated from a weighted average.

Seventy-eight sources were observed two or more times, and the scatters in their position estimates are consistent with the estimated uncertainties. Figure 2 displays a histogram of the ratio of the weighted rms scatter of the individual position coordinate estimates from the mean position of each multiply observed source, to the weighted rms estimate of position coordinate uncertainty. For the 78 multiply observed sources, the rms value of this ratio is 0.4 for right ascension and 0.5 for declination. Thus there is excellent agreement between the multiple observations of each source.

In addition to the 207 sources listed in Fig. 1, we also reobserved 22 sources whose positions were reported in Ref. 3. For these 22 sources, Fig. 3 shows a histogram of the absolute value of the difference in each position coordinate between the two separate observations divided by the RSS error of the position coordinate uncertainties. There is excellent agreement between the common sources. The calculated ratio has an RMS value of 1.3 for right ascension and 1.0 for declination. For one source, M104, an erroneous position was given in Ref. 3. The position for M104 cited in Fig. 1 should be used instead.

For 25 sources, we could compare our position estimates with other position estimates of better or similar accuracy (Refs. 7, 9, and 10). Figure 4 displays the ratio of the value of the difference between our source position estimate and the other catalog value to the RSS of the position uncertainties of both catalogs. For these 25 sources, the rms value of this ratio is 0.5 for right ascension and 0.9 for declination. For both right ascension and declination, the value of the ratio never exceeds 3.2. The bias offsets between our position estimates relative to those of the other catalogs for these 25 sources are 0ˢ.005 ±0ˢ.008 for right ascension and −0″.11 ±0″.06 for declination. Hence, these position estimate comparisons along with the multiple observation comparisons indicate our position uncertainty estimates are realistic.

Figure 5 shows histograms of the number of sources versus estimated position uncertainty for declination and right ascension, respectively. Estimated accuracies range from ∼0″.1 to 3″.8 in both right ascension and declination.

V. Conclusion

Positions for the milliarcsecond nuclei of 207 extragalactic sources have been determined to an accuracy of ~0.″1 to 3.″8 in both right ascension and declination. The reliability of the determined positions has been demonstrated by testing the repeatability of multiple observations on the same source and by comparing the results with other radio catalogs. Arcsecond positions have now been determined for 752 milliarcsecond nuclei.

Acknowledgments

We appreciate the assistance of G. Trager and J. Faulkner in data processing, and L. J. Skjerve, D. J. Spitzmesser, and the personnel of the Deep Space Network, Hartebeesthoek Radio Astronomy Observatory and Australian National Radio Astronomy Observatory at Parkes for aid in performing the observations.

References

1. Morabito, D. D., Preston, R. A., Faulkner, J., 1981, *Telecommunications and Data Acquisition Progress Reports, 42-66*, Sept. and Oct. 1981, Jet Propulsion Laboratory, Pasadena, Calif. p. 30.
2. Preston, R. A., Morabito, D. D., *NASA Conference Publication 2115*, Proceedings of a Conference held at MIT, Cambridge, Massachusetts, June 19-21, 1979.
3. Morabito, D. D., Preston, R. A., Slade, M. A., Jauncey, D. L., 1982, *Astronomical Journal, 87*, 517. Also published in *TDA Progress Report, 42-66*: September and October 1981, p. 37.
4. Fanselow, J. L., Thomas, J. B., Cohen, E. J., MacDoran, P. F., Melbourne, W. G., Mulhall, B. D., Purcell, G. H., Rogstad, D. H., Skjerve, L. J., Spitzmesser, D. J., Urech, J., Nicolson, G., 1979, *Time and the Earth's Rotation*, IAU Symp. 82, Reidel, p. 199.
5. Fanselow, J. L., Sovers, O. J., Thomas, J. B., Bletzacker, F. R., Kearns, T. J., Cohen, E. J., Purcell, G. H. Jr., Rogstad, D. H., Skjerve, L. J., Young, L. E., 1981, *Reference Coordinate Systems for Earth Dynamics*, pp. 351-357, D. Reidel Publishing Company.
6. Fricke, W. and Kopff, A. 1963, Veroeff. Astron. Rechen-Institut, Heidelberg, No. 10.
7. Waltman, E., Johnston, K. J., Spencer, J. H., Pauliny-Toth, I., Schraml, J., and Witzel, A., 1981, *Astron. Astrophys., 101*, 49.
8. McEwan, N. J., Browne, I. W. A., Crowther, J. H., 1975, *M.N.R.A.S., 80*, 1.
9. Ulvestad, J., Johnston, K., Perley, R., and Fomalont, E., 1981, *A.J., 86*, 1010.
10. Perley, R. A., 1982, *Astron. J. 87*, 859.

Table 1. Participating observatories

Location	Designation	Diameter, m	Baseline Length	
			Kilometers	Wavelengths
Tidbinbilla, Australia	DSS 43	64		
			10.6×10^3	8.1×10^7
Goldstone, California	DSS 14	64		
	DSS 13	26		
			8.4×10^3	6.4×10^7
Madrid, Spain	DSS 63	64		
Parkes, Australia	PK	64		
			9.7×10^3	7.5×10^7
Hartebeesthoek, So. Africa	HT	26		

Table 2. Experiment list

Experiment			Stations	Number of Position Calibrator Source Observations	Number of Source Positions Determined
Yr.	Mo.	Day			
78	05	30/31	13/43	4	27
78	06	29/30	13/43	8	10
80	02	01	14/43	4	7
80	02	27/28	14/43	3	10
80	03	03	13/63	8	25
80	03	12/13	13/63	13	26
80	03	28	13/63	4	25
80	04	24	PK/HT	9	17
80	04	25	PK/HT	11	14
80	04	25/26	PK/HT	10	14
80	06	19	14/43	6	9
81	05	08	14/43	3	1
82	02	14	43/HT	8	25
82	02	17	43/HT	3	33
82	02	19	43/HT	8	38
82	04	20	43/HT	6	43

Table 3. List of calibration sources (all positions from Ref. 5 unless otherwise noted)

P0003-00[2]
0016+73[1]
P0104-408
P0106+01
P0113-118
P0332-403
P0402-362
P0438-43
P0537-441
DA193
P0605-08
0615+82[1]
0716+71[1]
P0723-008[3]
P0727-11
P0743-006[3]
4C7107
OJ287
3C212[3]
4C39.25
P1104-445
P1127-14
P1144-379
1150+81[1]
3C273
3C274[3]
P1333-33[3]
3C309.1
DW1555+00
P1610-77[4]
3C345
NRA0530
P1741-038
1749+70[1]
1803+78[1]
OV-236
1928+73[1]
P1933-400
2007+77[1]
P2134+004
P2215+02[2]
OY-172.6
P2245-328
P2345-16

[1] Ref. 7
[2] Ref. 8
[3] Ref. 3
[4] D. L. Jauncey, et al., private communication

SOURCE NAME	RIGHT ASCENSION				DECLINATION			
	HR	MIN	SEC	ERROR	DEG	MIN	SEC	ERROR
P 0002-478	0	2	2.978	0.033	-47	53	1.81	0.30
0014+81	0	14	4.473	0.151	+81	18	28.71	0.22
0018+72	0	18	34.482	0.107	+72	56	3.93	0.24
P 0019-00	0	19	51.665	0.020	- 0	1	41.88	0.32
P 0022-423	0	22	15.413	0.020	-42	18	41.12	0.21
0027+70	0	27	17.026	0.083	+70	21	6.26	0.31
P 0034-01	0	34	30.555	0.020	- 1	25	37.78	0.36
P 0035-02	0	35	47.176	0.020	- 2	24	9.45	0.30
P 0036+03	0	36	44.176	0.020	+ 3	3	24.54	0.43
P 0047-579	0	47	48.238	0.064	-57	54	46.83	0.30
P 0055-01	0	55	1.572	0.020	- 1	39	39.40	0.34
P 0056-572	0	56	38.640	0.041	-57	15	22.07	0.30
P 0131-522	1	31	5.633	0.022	-52	15	26.40	0.17
P 0133-203	1	33	13.577	0.025	-20	24	4.01	0.44
0149+71	1	49	20.814	0.057	+71	0	20.76	0.24
B2 0149+33	1	49	40.027	0.024	+33	35	46.85	0.31
0153+74	1	53	4.339	0.054	+74	28	5.69	0.21
0159+72	1	59	13.059	0.057	+72	18	29.24	0.23
P 0202-76	2	2	1.005	0.422	-76	34	26.34	0.30
0205+72	2	5	26.907	0.050	+72	15	16.40	0.22
GC 0206+35	2	6	39.339	0.032	+35	33	41.36	0.40
P 0208-512	2	8	56.970	0.019	-51	15	7.53	0.17
0212+73	2	12	49.935	0.050	+73	35	40.15	0.21
P 0214-522	2	14	17.229	0.029	-52	14	5.49	0.21
P 0219-637	2	19	37.871	0.062	-63	43	59.84	0.21
P 0220-349	2	20	49.570	0.028	-34	55	4.67	0.35
GC 0223+34	2	23	9.735	0.027	+34	8	1.56	0.33
P 0226-038	2	26	21.984	0.021	- 3	50	57.25	0.42
P 0229-398	2	29	51.990	0.017	-39	49	0.15	0.18
P 0230-790	2	30	28.852	0.139	-79	1	0.77	0.21
P 0232-04	2	32	36.512	0.022	- 4	15	8.90	0.65
P 0235-618	2	35	37.782	0.034	-61	49	13.47	0.21
P 0244-452	2	44	4.579	0.023	-45	12	13.34	0.21
GC 0248+43	2	48	18.504	0.027	+43	2	56.98	0.34
P 0252-549	2	52	0.273	0.023	-54	54	2.27	0.17
P 0302-623	3	2	48.181	0.034	-62	23	3.71	0.21
P 0308-611	3	8	51.298	0.027	-61	9	58.24	0.17
P 0312-77	3	12	56.328	0.205	-77	3	0.15	0.30
P 0316-444	3	16	13.353	0.056	-44	25	11.24	0.30
DW 0326+27	3	26	56.037	0.026	+27	46	0.31	0.33
P 0332-403	3	32	25.244	0.026	-40	18	24.04	0.30
P 0334-546	3	34	36.057	0.029	-54	40	16.95	0.21
OE 367	3	40	14.804	0.029	+36	12	44.47	0.42
P 0355-66	3	55	27.892	0.095	-66	54	11.43	0.22
P 0355-483	3	55	52.505	0.027	-48	20	49.14	0.21
GC 0402+37	4	2	29.854	0.029	+37	55	26.85	0.42
0403+76	4	3	59.196	0.067	+76	48	52.75	0.25
P 0414-341	4	14	16.908	0.028	-34	10	26.06	0.37
3C 119	4	29	7.937	0.027	+41	32	8.24	0.31
P 0431-512	4	31	4.407	0.043	-51	15	42.11	0.23
P 0437-454	4	37	30.667	0.025	-45	28	12.26	0.22
P 0446-519	4	46	35.030	0.049	-51	56	13.75	0.23
P 0448-392	4	48	0.452	0.017	-39	16	15.74	0.18
P 0450-469	4	50	27.871	0.024	-46	58	16.45	0.21
P 0454+039	4	54	8.935	0.020	+ 3	56	14.71	0.41
P 0454-81	4	54	18.188	0.176	-81	5	54.13	0.21
P 0454-46	4	54	24.188	0.031	-46	20	38.47	0.23

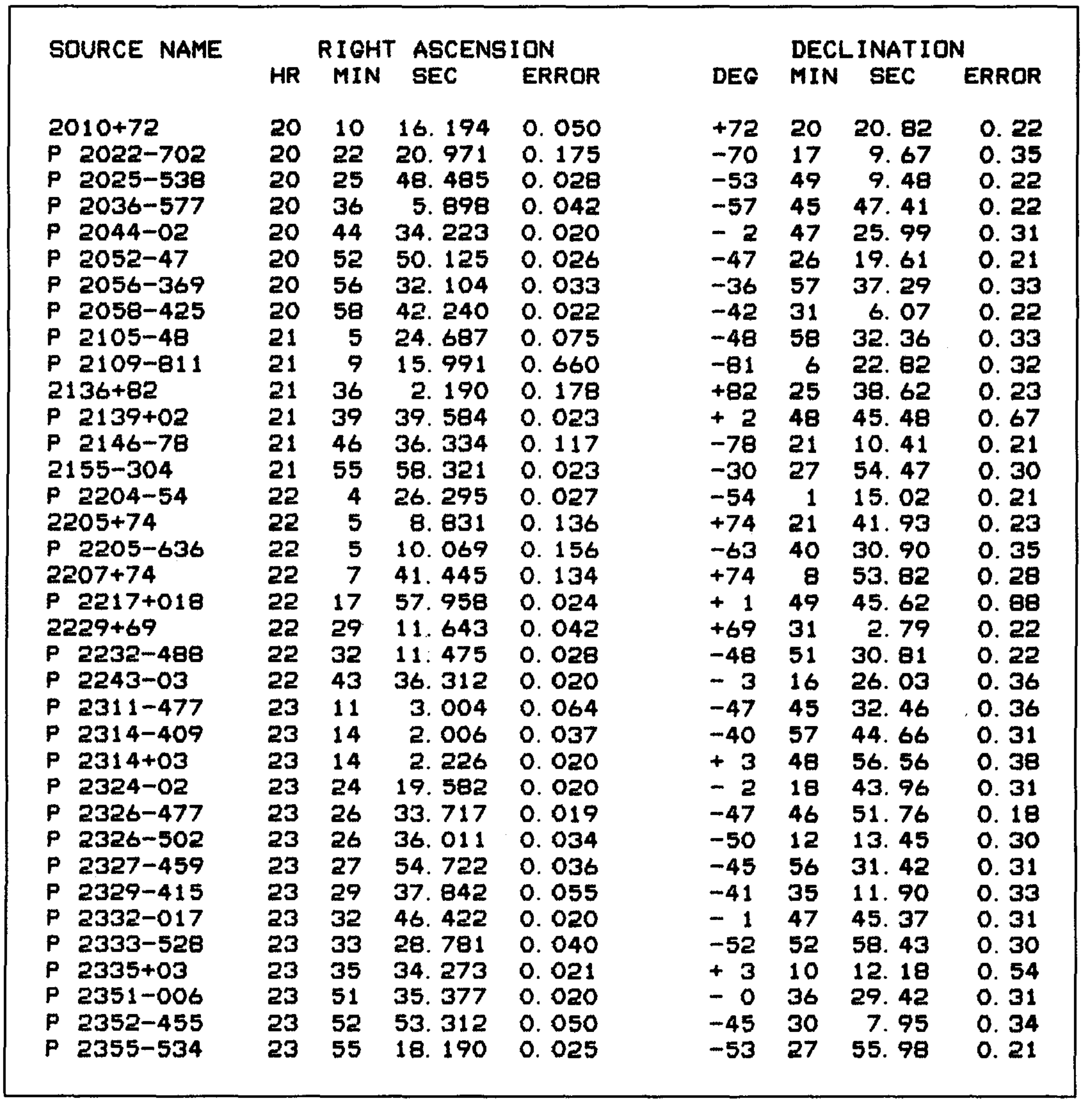
Fig. 1. Calculated positions and corresponding uncertainties of 207 sources

SOURCE NAME	RIGHT ASCENSION				DECLINATION			
	HR	MIN	SEC	ERROR	DEG	MIN	SEC	ERROR
0454+84	4	54	57. 212	0. 147	+84	27	53. 01	0. 21
P 0459+135	4	59	43. 842	0. 021	+13	33	56. 20	0. 31
P 0503-608	5	3	24. 298	0. 090	-60	53	56. 16	0. 31
P 0509+152	5	9	49. 674	0. 024	+15	13	49. 21	0. 41
P 0514-459	5	14	19. 327	0. 022	-45	59	58. 50	0. 18
P 0521-262	5	21	17. 209	0. 035	-26	16	52. 90	0. 55
P 0522-611	5	22	0. 426	0. 066	-61	10	41. 35	0. 31
P 0523-570	5	23	48. 149	0. 056	-57	1	27. 40	0. 23
P 0524-460	5	24	6. 001	0. 020	-46	0	28. 01	0. 18
0532+82	5	32	31. 216	0. 136	+82	36	53. 05	0. 23
P 0537-158	5	37	17. 184	0. 037	-15	52	5. 10	0. 35
P 0543-735	5	43	2. 748	0. 072	-73	33	31. 98	0. 18
0604+72	6	4	39. 237	0. 053	+72	49	27. 21	0. 23
P 0622-441	6	22	2. 686	0. 033	-44	11	22. 97	0. 22
P 0629-418	6	29	37. 684	0. 034	-41	52	14. 26	0. 30
0633+73	6	33	6. 438	0. 074	+73	27	35. 75	0. 31
P 0637-75	6	37	23. 422	0. 065	-75	13	37. 38	0. 15
0718+79	7	18	8. 892	0. 082	+79	17	22. 60	0. 22
P 0722+145	7	22	27. 228	0. 127	+14	31	7. 94	1. 81
P 0736+01	7	36	42. 553	0. 020	+ 1	43	59. 44	0. 31
0740+82	7	40	33. 222	0. 187	+82	49	24. 17	0. 22
P 0743-67	7	43	22. 189	0. 052	-67	19	9. 09	0. 31
P 0757-737	7	58	1. 046	0. 150	-73	44	57. 34	0. 30
0824+11	8	24	22. 311	0. 026	+11	2	19. 22	0. 37
0830+11	8	30	29. 914	0. 028	+11	33	52. 30	0. 37
P 0842-75	8	42	6. 608	0. 635	-75	29	19. 76	0. 40
P 0851+071	8	51	8. 451	0. 023	+ 7	6	11. 84	0. 33
P 0907-023	9	7	13. 129	0. 020	- 2	19	16. 43	0. 31
P 0915-213	9	15	10. 447	0. 026	-21	18	56. 88	0. 32
P 0925-203	9	25	33. 523	0. 021	-20	21	44. 65	0. 30
P 0931-114	9	31	8. 908	0. 020	-11	26	4. 59	0. 30
P 0932+02	9	32	43. 652	0. 021	+ 2	17	12. 21	0. 73
0950+74	9	50	4. 573	0. 113	+74	50	7. 67	0. 22
P 0957+00	9	57	43. 824	0. 020	+ 0	19	48. 94	0. 32
1003+83	10	3	25. 911	0. 307	+83	4	56. 52	0. 30
P 1020-103	10	20	4. 181	0. 083	-10	22	33. 40	0. 99
1023+747	10	23	13. 015	0. 121	+74	43	43. 43	0. 31
1027+74	10	27	13. 502	0. 424	+74	57	22. 48	0. 31
P 1030-357	10	30	52. 089	0. 035	-35	46	27. 46	0. 31
P 1032-199	10	32	37. 372	0. 021	-19	56	2. 19	0. 31
P 1034-058	10	34	16. 967	0. 049	- 5	50	16. 28	0. 95
P 1034-374	10	34	38. 242	0. 026	-37	28	39. 66	0. 30
1039+81	10	39	27. 818	0. 128	+81	10	23. 54	0. 21
1044+71	10	44	49. 741	0. 047	+71	59	26. 72	0. 21
P 1046-409	10	46	22. 645	0. 038	-40	58	7. 69	0. 33
1053+70	10	53	27. 729	0. 045	+70	27	47. 78	0. 21
1053+81	10	53	36. 302	0. 108	+81	30	35. 41	0. 17
P 1057-79	10	57	49. 877	0. 406	-79	47	47. 60	0. 31
1058+72	10	58	20. 119	0. 060	+72	41	44. 72	0. 22
P 1102-242	11	2	19. 825	0. 023	-24	15	13. 82	0. 30
P 1103-006	11	3	58. 359	0. 020	- 0	36	40. 84	0. 31
1104+72	11	4	17. 980	0. 080	+72	48	49. 95	0. 22
P 1105-680	11	5	17. 769	0. 139	-68	4	35. 70	0. 31
P 1106+023	11	6	11. 190	0. 020	+ 2	18	56. 23	0. 31
P 1110-217	11	10	21. 668	0. 024	-21	42	8. 72	0. 33
P 1116-46	11	16	6. 200	0. 032	-46	17	50. 01	0. 22
P 1120-274	11	20	28. 477	0. 026	-27	26	20. 21	0. 30

Fig. 1 (contd)

SOURCE NAME	RIGHT ASCENSION				DECLINATION			
	HR	MIN	SEC	ERROR	DEG	MIN	SEC	ERROR
P 1133-681	11	33	46.785	0.134	-68	10	29.41	0.30
P 1133-739	11	34	1.524	0.323	-73	59	8.91	0.31
P 1142-225	11	42	50.234	0.026	-22	33	51.75	0.33
P 1143-287	11	43	54.800	0.024	-28	42	37.96	0.31
P 1143-331	11	43	57.498	0.039	-33	12	2.60	0.45
P 1145-676	11	45	9.366	0.058	-67	37	1.66	0.35
P 1148-671	11	48	46.744	0.068	-67	11	29.38	0.22
P 1203-26	12	2	58.823	0.022	-26	17	22.59	0.30
P 1207-399	12	6	59.486	0.020	-39	59	31.49	0.18
P 1221-82	12	21	25.115	0.163	-82	56	33.54	0.31
1221+80	12	21	47.662	0.218	+80	56	41.03	0.21
P 1236-684	12	36	44.606	0.095	-68	29	2.18	0.31
M 104	12	37	23.378	0.013	-11	20	55.04	0.21
P 1240-294	12	40	30.049	0.028	-29	26	57.05	0.35
P 1250-330	12	50	14.899	0.025	-33	3	42.65	0.31
P 1251-71	12	51	40.129	0.063	-71	22	3.72	0.31
P 1256-229	12	56	27.621	0.023	-22	54	28.09	0.31
1305+80	13	5	22.143	0.191	+80	24	21.27	0.18
P 1347-218	13	47	28.024	0.024	-21	49	49.70	0.34
P 1349-439	13	49	52.576	0.029	-43	57	54.04	0.30
P 1349+027	13	49	58.378	0.020	+ 2	47	33.84	0.44
P 1351+021	13	51	18.912	0.020	+ 2	6	37.46	0.32
1357+76	13	57	42.177	0.133	+76	57	53.05	0.30
P 1406-267	14	6	58.442	0.024	-26	43	27.76	0.32
1436+76	14	36	4.574	0.317	+76	18	23.82	0.36
P 1438-347	14	38	20.331	0.027	-34	43	57.60	0.32
1448+76	14	48	56.492	0.078	+76	13	33.81	0.22
P 1451-400	14	51	20.602	0.018	-40	0	22.46	0.22
P 1509+022	15	9	43.717	0.020	+ 2	14	32.10	0.31
P 1549-79	15	49	28.382	0.211	-79	5	17.75	0.30
P 1602+01	16	2	12.955	0.020	+ 1	25	58.69	0.32
P 1602-00	16	2	21.909	0.020	- 0	10	57.17	0.32
P 1603+00	16	3	38.909	0.020	+ 0	8	30.12	0.36
P 1610-77	16	10	51.750	0.104	-77	9	52.60	0.30
1616+85	16	16	22.349	3.021	+85	9	26.04	0.42
P 1635-035	16	35	41.407	0.020	- 3	34	8.85	0.43
1637+82	16	37	56.847	0.332	+82	38	18.48	0.23
P 1706+006	17	6	11.695	0.020	+ 0	38	54.89	0.31
P 1718-649	17	18	46.160	0.022	-64	57	47.87	0.12
P 1719-729	17	19	52.046	0.089	-72	57	18.81	0.22
P 1732-598	17	32	2.807	0.046	-59	50	6.38	0.30
P 1733-56	17	33	24.401	0.045	-56	31	39.93	0.32
P 1758-651	17	58	25.634	0.037	-65	7	40.83	0.21
P 1800-709	18	0	36.223	0.099	-70	58	44.21	0.24
P 1815-554	18	15	35.184	0.044	-55	22	38.16	0.31
P 1823-455	18	23	31.139	0.034	-45	34	19.00	0.33
P 1831-711	18	31	41.335	0.078	-71	11	14.56	0.30
P 1853-534	18	52	59.134	0.042	-53	28	58.63	0.31
P 1925-610	19	25	40.692	0.034	-61	2	24.34	0.22
P 1929-457	19	29	8.055	0.034	-45	43	5.21	0.30
P 1935-692	19	35	11.833	0.116	-69	14	52.50	0.39
P 1936-623	19	36	52.756	0.067	-62	18	21.43	0.30
P 1941-554	19	41	23.321	0.038	-55	28	6.09	0.23
1946+70	19	46	12.035	0.072	+70	48	21.66	0.25
P 2004-447	20	4	25.155	0.031	-44	43	28.51	0.30
P 2005-489	20	5	46.605	0.034	-48	58	43.55	0.30
2007+77	20	7	20.428	0.094	+77	43	57.95	0.30

Fig. 1 (contd)



SOURCE NAME	RIGHT ASCENSION				DECLINATION			
	HR	MIN	SEC	ERROR	DEG	MIN	SEC	ERROR
2010+72	20	10	16.194	0.050	+72	20	20.82	0.22
P 2022-702	20	22	20.971	0.175	-70	17	9.67	0.35
P 2025-538	20	25	48.485	0.028	-53	49	9.48	0.22
P 2036-577	20	36	5.898	0.042	-57	45	47.41	0.22
P 2044-02	20	44	34.223	0.020	- 2	47	25.99	0.31
P 2052-47	20	52	50.125	0.026	-47	26	19.61	0.21
P 2056-369	20	56	32.104	0.033	-36	57	37.29	0.33
P 2058-425	20	58	42.240	0.022	-42	31	6.07	0.22
P 2105-48	21	5	24.687	0.075	-48	58	32.36	0.33
P 2109-811	21	9	15.991	0.660	-81	6	22.82	0.32
2136+82	21	36	2.190	0.178	+82	25	38.62	0.23
P 2139+02	21	39	39.584	0.023	+ 2	48	45.48	0.67
P 2146-78	21	46	36.334	0.117	-78	21	10.41	0.21
2155-304	21	55	58.321	0.023	-30	27	54.47	0.30
P 2204-54	22	4	26.295	0.027	-54	1	15.02	0.21
2205+74	22	5	8.831	0.136	+74	21	41.93	0.23
P 2205-636	22	5	10.069	0.156	-63	40	30.90	0.35
2207+74	22	7	41.445	0.134	+74	8	53.82	0.28
P 2217+018	22	17	57.958	0.024	+ 1	49	45.62	0.88
2229+69	22	29	11.643	0.042	+69	31	2.79	0.22
P 2232-488	22	32	11.475	0.028	-48	51	30.81	0.22
P 2243-03	22	43	36.312	0.020	- 3	16	26.03	0.36
P 2311-477	23	11	3.004	0.064	-47	45	32.46	0.36
P 2314-409	23	14	2.006	0.037	-40	57	44.66	0.31
P 2314+03	23	14	2.226	0.020	+ 3	48	56.56	0.38
P 2324-02	23	24	19.582	0.020	- 2	18	43.96	0.31
P 2326-477	23	26	33.717	0.019	-47	46	51.76	0.18
P 2326-502	23	26	36.011	0.034	-50	12	13.45	0.30
P 2327-459	23	27	54.722	0.036	-45	56	31.42	0.31
P 2329-415	23	29	37.842	0.055	-41	35	11.90	0.33
P 2332-017	23	32	46.422	0.020	- 1	47	45.37	0.31
P 2333-528	23	33	28.781	0.040	-52	52	58.43	0.30
P 2335+03	23	35	34.273	0.021	+ 3	10	12.18	0.54
P 2351-006	23	51	35.377	0.020	- 0	36	29.42	0.31
P 2352-455	23	52	53.312	0.050	-45	30	7.95	0.34
P 2355-534	23	55	18.190	0.025	-53	27	55.98	0.21

Fig. 1 (contd)

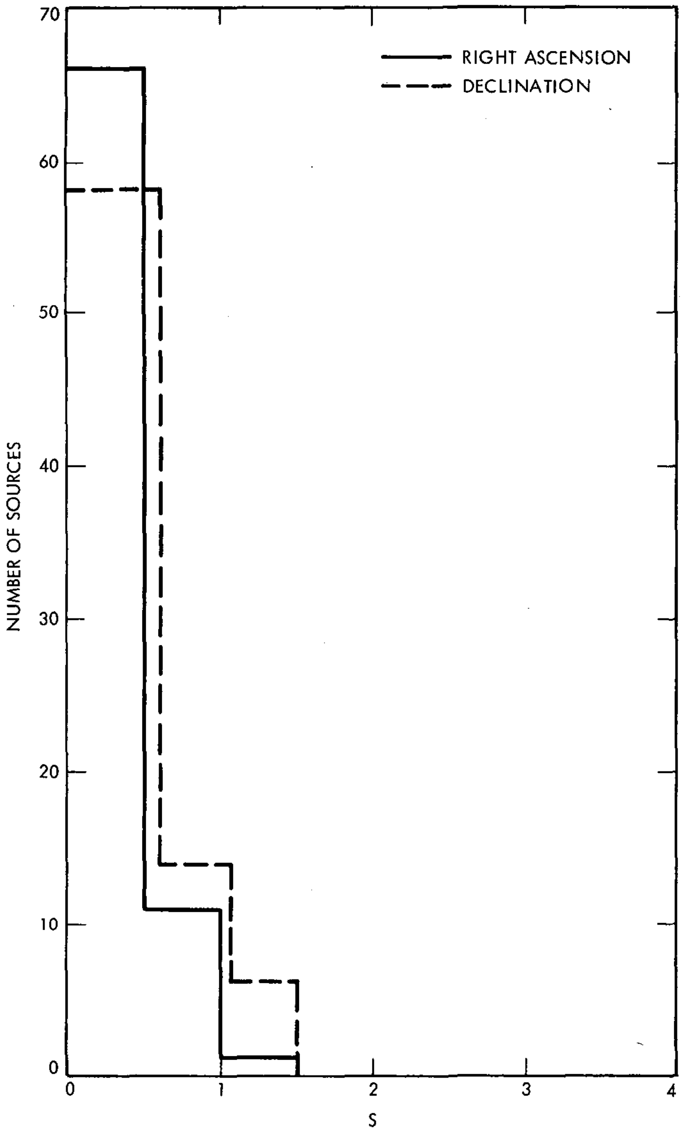


Fig. 2. Comparison of position estimates for 78 multiply observed sources. For each source, *S* is the ratio of the weighted rms scatter of individual position coordinate estimates about the mean value to the weighted rms estimate of the uncertainty in that position coordinate.

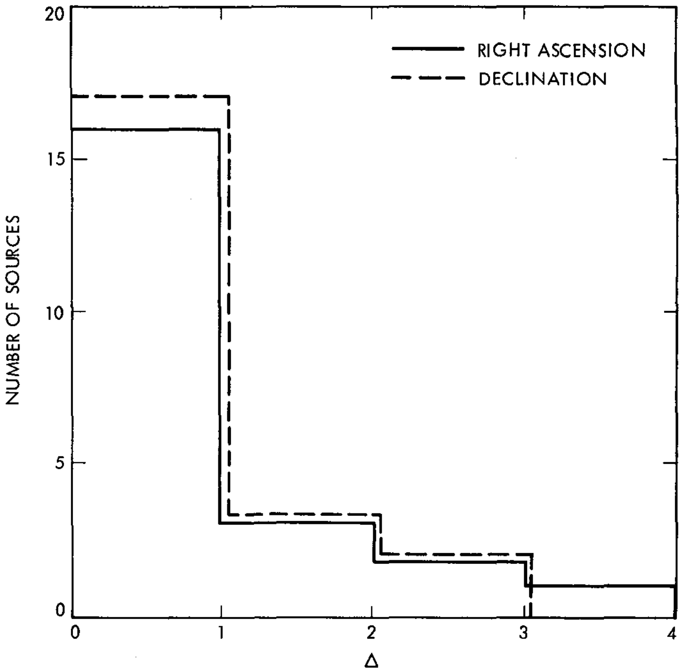


Fig. 3. Comparison of position estimates with 22 sources common with Ref. 3. For each source, Δ is the ratio of the absolute value of the difference between the source coordinate estimate in this set of data and the source coordinate estimate cited in Ref. 3 to the weighted rms estimate of the uncertainty in that position coordinate.

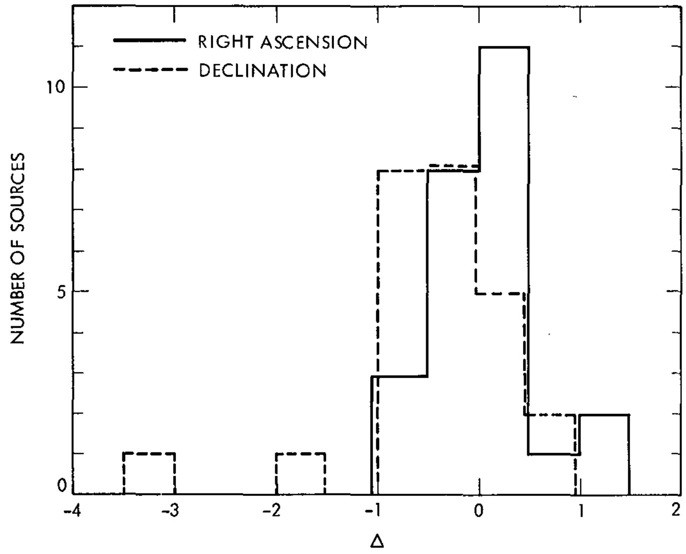


Fig. 4. Comparison of position estimates with other radio positional catalogs. For each source, Δ is the ratio of the difference between our source coordinate estimate and the other catalog value to the RSS of the uncertainties of both catalogs.

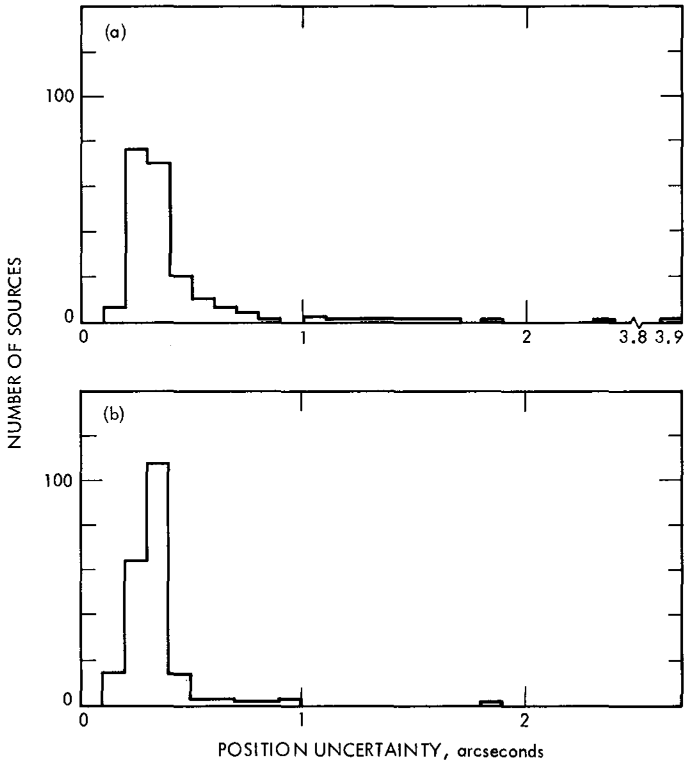


Fig. 5. Histogram of number of sources versus estimated position uncertainty: (a) right ascension; (b) declination

Submilliarcsecond VLBI Observations of the Close Pair GC 1342+662 and GC 1342+663

D. D. Morabito
Tracking Systems and Applications Section

Differential VLBI has been simultaneously performed on the source pair GC 1342+662 and GC 1342+663 (4.4-arcminute separation) using S-band on the Goldstone/Madrid baseline. These measurements were acquired on two separate observing sessions: 30 December 1982 and 14 May 1983. The change in separation of GC 1342+662 relative to GC 1342+663 between the two epochs was 0.03 ±0.08 milliarcsecond. The differences of the relative position measurements between epochs of GC 1342+662 relative to GC 1342+663 were −0.29 ±0.05 milliarcsecond in right ascension and 0.14 ±0.09 milliarcsecond in declination. These measurements demonstrate submilliarcsecond accuracy and repeatability. The discrepancies outside of the formal uncertainties could be attributed to the intrinsic properties of the sources such as structure and to a lesser probability, proper motion. These discrepancies could also be attributed to excursions in UT1-UTC of about four times the quoted BIH uncertainty.

I. Introduction

VLBI reference frames of several milliarcsecond precision are being used by the Deep Space Network for spacecraft navigation (Refs. 1, 2, and 3). Quasars from these reference frames have served as reference beacons in differential VLBI experiments with the Voyager spacecraft, the Pioneer Venus Orbiter (Ref. 3), and the Viking Orbiter (XX Newhall, private communication).

The potential limiting accuracy attainable for differential VLBI positional measurements, using these reference frames, can be reached by simultaneously observing sources with small angular separations. These close pairs of sources must have angular separations smaller than the beamwidth of each antenna, and therefore can be simultaneously tracked by both antennas of a VLBI baseline, greatly enhancing the cancellation of measurement errors. By observing several close pairs of radio sources over a several-year period, limits on reference frame stability can be determined.

Observations and results demonstrating the submilliarcsecond accuracy of this technique at 2.3 GHz using the DSN baseline, Goldstone/Madrid, on the close pair GC 1342+662 and GC 1342+663 are presented. These two sources have been identified as quasars (Ref. 4), and have originally been identified as VLBI sources in the DSN full-sky survey program (Ref. 5). Total-flux densities measured during April 1983 indicate that both of these sources have flat spectra (R. Perley, private communication).

II. Observations

The close pair GC 1342+662 and GC 1342+663 was observed at 2.29 GHz with the Goldstone/Madrid baseline. It was observed by DSS 14 (64 m) and DSS 63 (64 m) on 30 December 1982 for 6 hours and 40 minutes, and by DSS 13 (26 m) and DSS 63 (64 m) on 14 May 1983 for 6 hours. The Goldstone/Madrid baseline has a length of 8400 km and a fringe spacing of 3 milliarcseconds. The observations were performed with right circular polarization. The receiver chain consisted of an S-band traveling wave maser followed by a special phase-stable S-band VLBI receiver, which converted the RF signal to an IF of 50 MHz. The NRAO Mark II VLBI recording system then recorded a 1.8-MHz data bandwidth by digitally sampling at a 4-Mbits/s rate (Ref. 6). Digital sampling and phase stability of the receiver chain were controlled by hydrogen maser atomic clocks. The centroid position of the source pair was used to point the antenna. X-band observations were not performed since the 4.4-arcminute angular separation of this source pair exceeded the half-power beamwidth of the limiting antenna (2 arcminutes), thus making simultaneous observations useless.

The data were correlated on the Caltech/JPL Mark II correlator. The postprocessing software yielded 60-second points of fringe phase for each source. The phases were then input to the JPL Orbit Determination Program (Refs. 7, 8, and 9), where an accurate phase model was applied. The phases were then corrected for cycle ambiguities and differenced. This difference phase was input to a least-squares algorithm where an integral number of cycles was determined and then added to each of the difference phases. The next iteration of the least-squares algorithm was a two-parameter fit that yielded the relative position offsets in right ascension and declination.

III. Results

Figure 1 shows the postfit difference-phase residuals for the 30 December 1982 experiment, which have an rms scatter of 0.028 cycles (90 μarcseconds). Figure 2 shows the postfit difference-phase residuals for the 14 May 1983 experiment, which have an rms scatter of 0.026 cycles (80 μarcseconds). The scatters reported here are in agreement with predicted values and are dominated by the solar plasma. The scatter on the individual 60-second integrated phase points is about 0.003 cycles and is consistent with expected noise levels due to system noise and ionosphere over these time scales.

The correlated amplitudes were converted to source strengths using measured values of system temperatures and antenna efficiency parameters. The resulting correlated flux densities at 2.3 GHz are shown in Fig. 3 for GC 1342+663 and in Fig. 4 for GC 1342+662. Peak correlated flux densities on 30 December 1982 were 0.35 Jy for GC 1342+662 and 0.75 Jy for GC 1342+663, and on 14 May 1983 they were 0.35 Jy for GC 1342+662 and 0.60 Jy for GC 1342+663. The correlated flux density behavior for GC 1342+663 is essentially flat (see Fig. 3) indicating a lack of detectable structure over the observed hour-angle range. There is, however, a significant variation of correlated flux density over the interferometric hour angle for GC 1342+662 (see Fig. 4) indicating complex structure over the milliarcsecond resolution beam of the interferometer. This variation is consistent with a model of two point sources with equal flux densities (0.175 Jy each) separated by 1.1 milliarcseconds and aligned along a $\approx$29-degree position angle (defined north through east). The position angle is coincidentally aligned with the position angle between the two sources GC 1342+662 and GC 1342+663, suggesting the possibility of an intrinsic relationship between the sources.

The measured position in arclength of GC 1342+662 relative to GC 1342+663 referred to epoch 1950.0 is given in Table 1 for both experiments. The arclength differences of these relative position measurements between experiments were -0.29 ±0.05 milliarcseconds in right ascension and 0.14 ±0.09 milliarcseconds in declination. The results agree to within 5.8 and 1.5 times the rss of the individual standard deviations for right ascension and declination respectively. Marcaide and Shapiro (Ref. 10) had similar few-sigma discrepancies between epochs of their relative position measurements of 1038+528A and 1038+528B. The change in separation of the GC 1342+662 relative to GC 1342+663 between the two epochs was 0.03 ±0.08 milliarcsecond, implying that the above discrepancies are effectively a rotation in right ascension and declination space. Because most of the discrepancy is in right ascension, timing errors are immediately suspect. However, such errors would require deviations from BIH values of UT1-UTC of about four times the quoted uncertainty (Ref. 11).

Source structure is the most likely cause of the measured discrepancy in the differential position measurements between epochs for GC 1342+662 and GC 1342+663. Proper motion is a possible cause for the discrepancy, but not very probable assuming current cosmological assumptions. To properly understand and distinguish between these effects, more observations need to be performed over a several-year time period.

Marcaide and Shapiro (Ref. 10), Shapiro et al. (Ref. 11), and Gorenstein et al. (Ref. 12) have performed similar experiments on other close pairs. Table 2 displays a summary of these experiments along with the one discussed in this paper.

IV. Conclusion

The improved accuracy and repeatability of simultaneous differential VLBI using the close pair GC 1342+662 and GC 1342+663 with the Goldstone/Madrid baseline has been demonstrated on two separate observing sessions. The change in separation of GC 1342+662 relative to GC 1342+663 between the two epochs was 0.03 ±0.08 milliarcsecond. The differences of the relative position measurements of GC 1342+662 relative to GC 1342+663 between the two epochs were -0.29 ±0.05 milliarcseconds in right ascension and 0.14 ±0.09 milliarcseconds in declination. Discrepancies between the relative measurements outside the formal uncertainties could be attributed to either source-structure effects or proper motion. Further monitoring of this close pair along with other close pairs will be useful for testing and placing limits on reference frame stability.

Acknowledgments

I would like to thank D. Trask, G. H. Purcell, A. E. Niell, R. A. Preston, and T. P. Yunck for many greatly appreciated discussions. I would also like to thank X X Newhall for running the ODP software, and L. Skjerve, B. Johnson, E. Lobdell, and S. Harris, along with the staff of DSS 63, for performing the observations.

References

1. Melbourne, W. G., Curkendall, D. W., "Radio Metric Direction Finding: A New Approach to Deep Space Navigation," AAS/AIAA Astrodynamics Specialist Conference Paper presented at Jackson Hole, Wyoming, Sept. 7-9, 1977.
2. Brunn, D. L., Preston, R. A., Wu, S. C., Siegel, H. L., Brown, D. S., "ΔVLBI Spacecraft Tracking System Demonstration: Part I. Design and Planning," *DSN Progress Report 42-45, March and April 1978*, pp. 111-132.
3. Esposito, P. B., Donivan, F. F., Finley, S. G., Newhall, X X, Smith, C. B., Wu, S. C., "Narrowband Differential Interferometry Applied to Pioneer Venus Orbiter," paper presented at AAS/AIAA Astrodynamics Specialist Conference held at Lake Placid, New York, Aug. 22-25, 1983.
4. Kuhr, H., Nauber, U., Pauliny-Toth, I. I. K., Witzel, A., "A Catalogue of Radio Sources," Max-Planck-Institut fur Radioastronomie, Bonn, Germany, Preprint 55, August 1979.
5. Morabito, D. D., Preston, R. A., Slade, M. A., and Jauncey, D. L., "Arcsecond Positions for Milliarcsecond VLBI Nuclei of Extragalactic Radio Sources. I. 546 Sources," *Astronomical Journal*, Vol. 87, No. 3, pp. 517-527, March 1982.
6. Clark, B. G., "The NRAO Tape-Recorder Interferometer System," *Proceedings of the IEEE*, Vol. 61, p. 1242, 1963.
7. Moyer, T. D., *Mathematical Formulation of the Double-Precision Orbit Determination Program (DPODP)*, TR 32-1527, Jet Propulsion Laboratory, Pasadena, California, May 15, 1971.
8. Moyer, T. D., "Transformation From Proper Time on Earth to Coordinate Time in Solar System Barycentric Space-Time Frame of Reference: Part I," *Celestial Mechanics*, Vol. 23, p. 33. D. Reidel Publishing Company, 1981.

9. Moyer, T. D., "Transformation From Proper Time on Earth to Coordinate Time in the Solar System Barycentric Space-Time Frame of Reference: Part 2," *Celestial Mechanics*, Vol. 23, p. 57. D. Reidel Publishing Company, 1981.

10. Marcaide, J. M., and Shapiro, I. I., "High Precision Astronomy via Very-Long-Baseline Radio Interferometry: Estimate of the Angular Separation Between the Quasars 1038+528A and B," *Astronomical Journal*, Vol. 88, p. 1133, 1983.

11. Shapiro, I. I., Wittels, J. J., Counselman, C. C., III, Robertson, D. S., Whitney, A. R., Hinteregger, H. F., Knight, C. A., Rogers, A. E. E., Clark, T. A., Hutton, L. K., Niell, A. E., "Submilliarcsecond Astrometery via VLBI. I. Relative Position of the Radio Sources 3C345 and NRAO 512," *Astronomical Journal*, Vol. 84, p. 1459, 1979.

12. Gorenstein, M. V., Shapiro, I. I., Cohen, N. L., Falco, E., Kassim, N., Rogers, A. E. E., Whitney, A. R., Preston, R. A., Rius, A., "VLBI Observations of the "Twin Quasar" 0957+56A and B," *B.A.A.S.*, Vol. 12, p. 498, 1980.

Table 1. Relative position measurements

Source	Epoch	Relative right ascension,[a] arcseconds	Relative declination, arcseconds	Total separation, arcseconds
GC 1342+662	30 Dec 1982	−139.58033 ±0.00004	−225.17070 ±0.00006	264.83216 ±0.00006
GC 1342+662	14 May 1983	−139.58062 ±0.00003	−225.17056 ±0.00007	264.83219 ±0.00006

[a] In arclength referred to declination of GC 1342+662.

Table 2. Comparison with other investigations

Investigators	Source pair	Angular separation, arcminutes	Measurement accuracy, μarc
Ref. 11	3C 345 NRAO 512	30	500
Ref. 12	0957+561A 0957+561B	0.1	50
Ref. 10	1038+528A 1038+528B	0.6	10
This paper	GC 1342+662 GC 1342+663	4.4	60

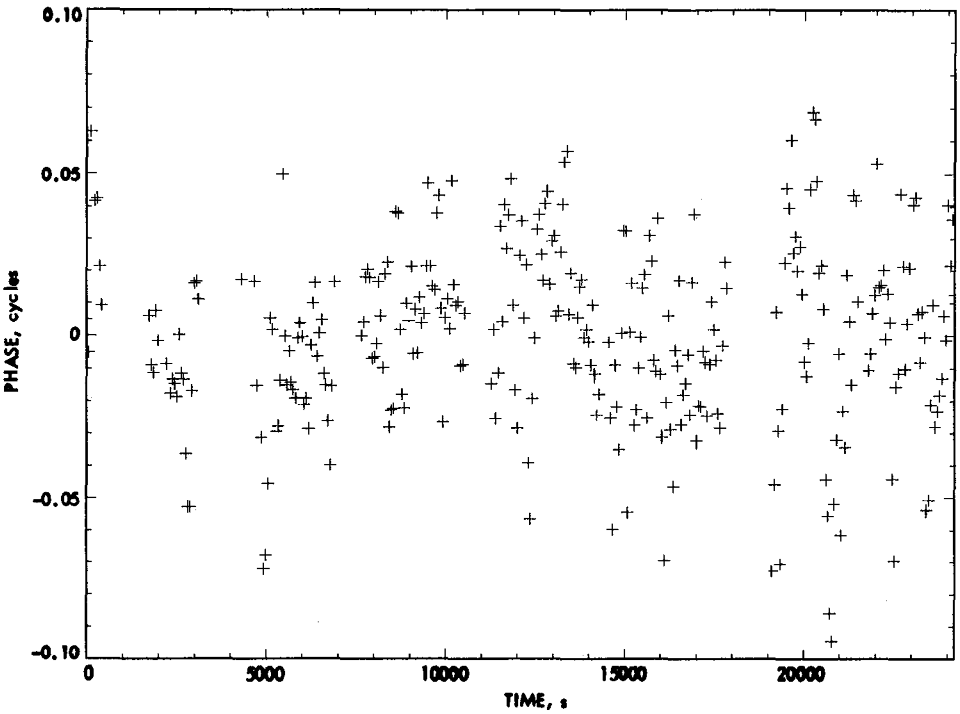


Fig. 1. Postfit residual difference phase, 30 December 1982, baseline DSS 14/DSS 63

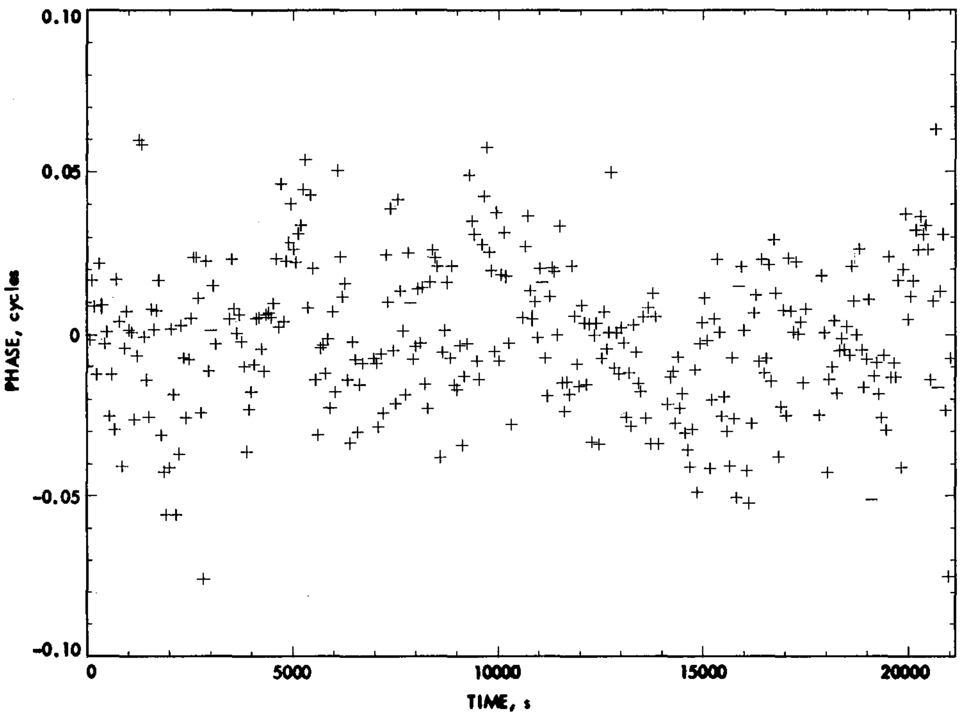


Fig. 2. Postfit residual difference phase, 14 May 1983, baseline DSS 13/DSS 63

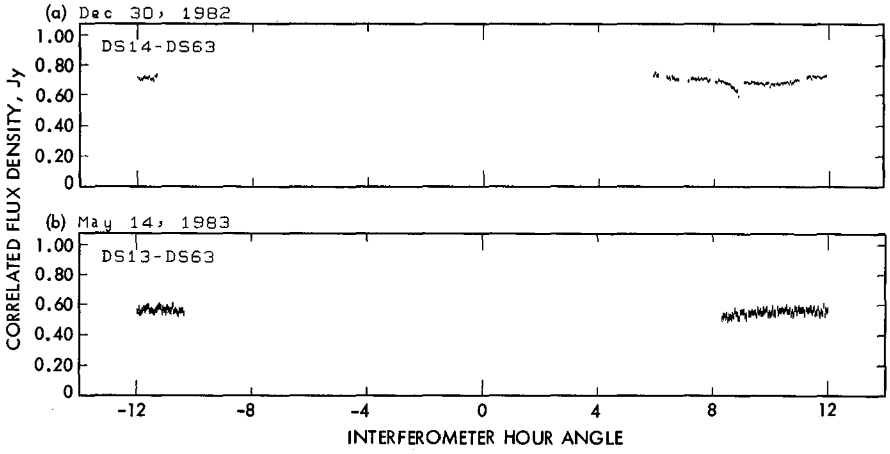


Fig. 3. Correlated flux density vs interferometer hour angle for GC 1342+663: (a) first experiment; (b) second experiment

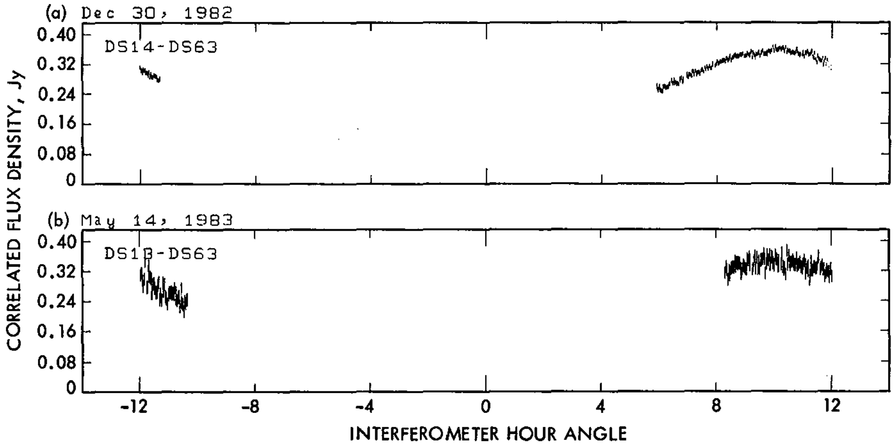


Fig. 4. Correlated flux density vs interferometer hour angle for GC 1342+662: (a) first experiment; (b) second experiment

Magnetically Enhanced Hydrogen Gas Dissociator: A Progress Report

G. Lee, T. K. Tucker, and L. Maleki
Communications Systems Research Section

Application of a sensitive hydrogen gas detector has led to the observation of a significant increase in the efficiency of atom production of a magnetically enhanced dissociator. An increase of about 7000% in the production of atoms in the hyperfine state desired for maser operation is observed at certain values of the applied magnetic field. A preliminary test of the dissociator in the maser has led to an observed increase of 6.18 dB in the maser power.

I. Introduction

The development of the magnetically enhanced hydrogen gas dissociator was initiated in response to a need for improving the operational life of the hydrogen maser. An improvement in the efficiency of the dissociator would result in reduced gas consumption, extending the operational life of vacuum pumps. This would reduce the relatively large number of maser failures which result from the breakdown of Vacion pumps, as often encountered in the field.

Previous investigations involving the application of an axial dc magnetic field on the rf excited dissociator had led to encouraging results (Refs. 1 and 2). The results of these investigations were based on the determination of the flux of the atomic hydrogen effusing from the dissociator with and without the applied magnetic field. These measurements, however, could only yield relative values for the atomic flux; the absolute value of the hydrogen atom flux could not be obtained because of the nature of the measuring technique (Ref. 2).

In the present work we report on our observation of the atomic hydrogen flux as measured directly with a hydrogen atom detector. The use of this detector permitted a further investigation of the dependency of the atomic flux on the strength of the dc magnetic field. The latter results, together with the details of the hydrogen atom detector are subjects of this article.

Section II of this article presents the details of the hydrogen atom detector and Section III provides the result of investigations using the detector. The study will be summarized in Section IV followed by recommendations for future investigations.

II. The Hydrogen Atom Detector

An efficient detection of atomic hydrogen may be accomplished through measurement of the heat released when two hydrogen atoms combine to form a molecule. The amount of heat produced in the reaction $H + H \rightarrow H_2$ is large enough (4.38 eV/reaction) for detection via the measurement of change of resistance of a suitable wire. It can be shown that for a length $2a$ of wire immersed in hydrogen the change in the resistance ΔR is given by:

$$\Delta R = \frac{(R\alpha n\gamma EaD)}{(\pi kd)}$$

Here α is the temperature coefficient of the wire, k is the thermal conductivity, n is the number of atoms per unit area per unit time, γ is the probability of recombination for H on the particular material of the wire, $2E$ is the heat of recombination, d is the diameter of the wire, and D is the atomic beam diameter.

In our studies the expected flux of atoms out of the dissociator is of the order of 10^{15} atoms per second. The atom detector therefore had to be designed with a sensitivity high enough to measure such a flux accurately and efficiently. The apparatus, designed with the above constraint, consisted of two short segments of a thin (2.5 μm in diameter) platinum wire arranged in an ordinary dc bridge circuit (see Fig. 1). The platinum wire was prepared by etching the surface of a Wollaston wire with nitric acid to expose the thin platinum core.

Only one of the segments was exposed to the hydrogen beam, so the difference in the voltage across this segment and that across the second segment could be amplified with a differential amplifier. In this way the detector was sensitive to a hydrogen flux as small as 10^{11}. The detector and its electronic elements were carefully shielded from the rf power used to excite the dissociator. This eliminated a relatively large noise problem which was initially encountered with the detector.

Since the probability of the recombination of H on platinum changes with surface contamination, provisions were made in the detector to periodically clean the surface of the wire through passage of a relatively large (2 mA) current. Thus the sensitivity of the detector was kept at a constant value throughout each run.

The platinum wire detector was placed in a stainless steel vacuum chamber, evacuated with a Vacion pump. The detector was placed on a plate attached to a rotatable feedthrough which could be used to move the plate and the detector in the horizontal direction. The rf excited source was placed external to the vacuum on a flange so that the holes in its exit collimator defined a vertical beam of atoms aligned to intersect the platinum wire. The source, rf exciter, and solenoid providing the dc field are identical to those described previously (Ref. 2).

III. Results and Measurements

As a first test of the hydrogen atom detector we attempted to replicate the previously reported results obtained with the magnetically enhanced dissociator. Under the normal conditions of source pressure and rf power the application of the dc magnetic field resulted in an increase in the source efficiency which agreed closely to the values reported in Refs. 1 and 2.

The increased efficiency of the atom production with the magnetically enhanced dissociator was large enough (up to 70% for certain values of pressure and source power) that an optimization of the beam optics for application in the maser seemed appropriate. We therefore modified the dissociator test apparatus to support a state selecting hexapole magnet at the exit end of the source. This magnet selects atoms of hydrogen that are in the hyperfine ground state having $F = 1$ and $m_F = 1$ and 0 and focuses them on the entrance hole of the storage bulb in the maser. The remainder of the atoms in the $F = 0$ state are defocused and therefore do not enter the storage bulb. This selection is necessary because all atoms that are not in the $F = 1, m_f = 1$ state do not participate in the maser oscillation, and yet they contribute to line broadening through spin exchange collisions.

The observed atom production efficiency of the magnetically enhanced dissociator with the state selecting hexapole magnet in line was quite unexpected. There appeared to be a very significant increase in the production of atoms focused by the state selector. While the flux of atoms which are focused by the hexapole magnet is typically of the order of 10^{13} atoms per second, the application of the dc magnetic field on the source increased the flux of focused atoms by as much as 7000% for a particular value of the applied field. Beyond this, a plot of atomic flux vs. magnetic field revealed pronounced structures which in certain cases implied a decrease to zero for the flux of focused atoms. This result is presented in Fig. 1.

As indicated in Fig. 2, there are a number of features in the plot of flux vs. field. There are two broader peaks at about 150 and 500 Gauss, and a rather sharp peak at about 230 Gauss. A number of valleys are also present including one at about 30 Gauss, which indicates a decrease in the flux of the atoms to zero value. These results were checked for reproducibility many times and were verified each time. They did however indicate changes in the amplitude of the peaks as parameters of rf power and hydrogen pressure were changed.

In view of the remarkable observation presented in Fig. 2, we tested the measuring apparatus to ensure the reliability of results and absence of artifacts. We also examined possible changes in the focusing properties of the hexapole magnet as a result of the presence of the applied dc field. In all cases we reached the conclusion that the results obtained were not due to any artifacts associated with the technique of measurement or the measuring apparatus.

The observed results then clearly indicate that for certain values of the dc magnetic field the total efficiency of atom production of the source is increased. But more importantly, the results indicate a much larger increase, at certain field values, in the production of those atoms in the desired hyperfine state.

This last contention was tested by retrofitting a maser with the magnetically enhanced dissociator and measuring the maser output power. Because of the physical requirement that the solenoid producing the dc field be outside the vacuum chamber, the dissociator had to be placed external to the vacuum chamber of the maser. This produced an increase in the path length of hydrogen atoms by a factor of nearly two, as compared to the path length of atoms when the dissociator is in its usual configuration inside the vacuum chamber.

The result of this test indicated that the maser oscillation ceased when the dc field was at a value which indicated zero atom production in the test bed (about 30 Gauss). As the field increased the maser power increased and reached a level 6.18 dB above its value without the magnetic field for fields larger than about 120 Gauss. The 6.18 dB increase in the power level, however, remained constant for all field values above 120 Gauss.

While this observation verified that the efficiency of atom production in the desired hyperfine state is significantly increased with the application of a dc field, it failed to establish a clear correlation between the observed peaks of Fig. 1 and the maser power. We believe that a clear understanding of this observation will be possible only when modifications are made to ensure that the path length of the atoms in shortened to be equivalent to the length in a normal maser configuration.

At this point it may be appropriate to discuss the possible causes of the behavior of H-atom flux with applied dc field when the hexapole magnet is placed in the beam path. A possible explanation of the observed result is the preferential pumping of the levels of the ground and excited electronic states of hydrogen by the rf power and the cyclotron radiation of electrons circulating in the applied magnetic field. For example, the peak at 230 Gauss could be attributed to pumping of the $2S$, $F = 1$, $m_f = -1$ state, which is forbidden to make a transition to the desired ground state, to the $2P$, $F = 1$, $m_f = 1$ state, which is allowed to make the transition. While we have similar explanations for other features in the spectrum depicted in Fig. 2, we believe that a totally conclusive explanation should await further studies with the source in the modified test chamber.

IV. Summary

In this article we have described some details of a hydrogen atom detector which has enabled a comprehensive investigation of the atom production of the magnetically enhanced dissociator. Our observations have revealed that while a significant increase in the production of atoms is obtained, the magnetically enhanced dissociator can increase at a greater proportion the number of atoms in the desired hyperfine state. This observation was further supported when a large (415%) increase was produced in the power output of the maser, which was operated in a test configuration with the magnetically enhanced dissociator. Finally, a preliminary explanation was put forth to describe the physical processes which can lead to the features observed in the spectrum of atomic flux vs. the magnetic field.

While the above results are to be considered quite preliminary, they do imply rather significant contribution to H-maser technology. As discussed in previous works (Refs. 1 and 2), the use of the axially applied magnetic field with the dissociator can enable maser operation at a significantly reduced hydrogen throughput. This will increase the operational life of the vacuum pumps without degrading the maser power output. The observation of the significant increase of the atoms in the desired hyperfine state further implies that for a fixed output power of maser (and therefore a fixed number of desired atoms in the storage bulb)' the magnetically enhanced dissociator can lead to significantly reduced line broadening due to spin exchange collisions. This may be brought about since the applied field can evidently increase the ratio of "desired atoms" to "unwanted atoms" in the storage bulb of the maser cavity. Therefore significantly higher line *Q*'s should be expected with the use of the magnetically enhanced dissociator, when all other parameters are kept fixed.

The encouraging results of this investigation have led us to believe that the above investigations should continue to fully determine the ultimate influence of the enhanced source on the operation of the H-maser. Currently further studies are underway to shorten the optical path of the atoms to the normal length as well as to pursue investigations with the test chamber.

References

1. Maleki, L., and Tucker, T. K., "The Development of a Magnetically Enhanced Hydrogen Gas Dissociator." *TDA Progress Report, 42-67,* November-December 1981, pp. 24-27, Jet Propulsion Laboratory, Pasadena, California.
2. Maleki, L., "The Development of a Magnetically Enhanced Hydrogen Gas Dissociator." *Proceedings of the 36th Annual Symposium on Frequency Control,* 2-4 June 1982 Philadelphia, Pennsylvania.

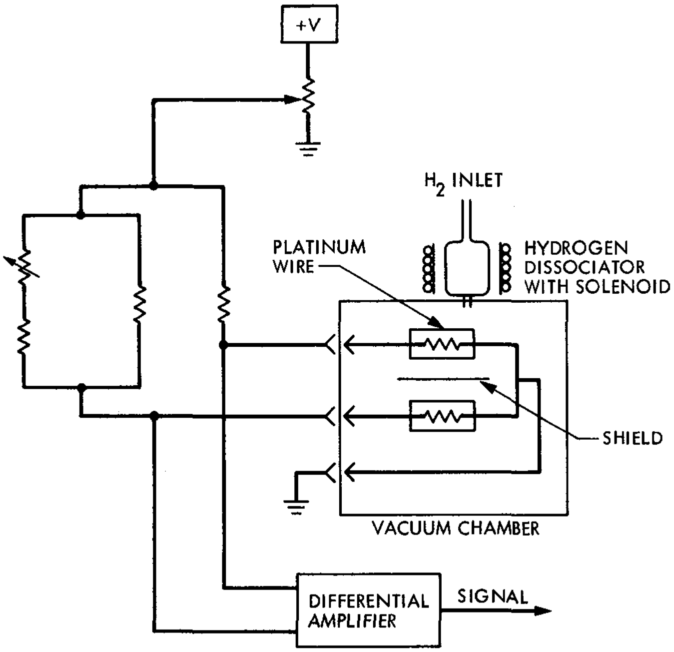


Fig. 1. Block diagram of the hydrogen detector

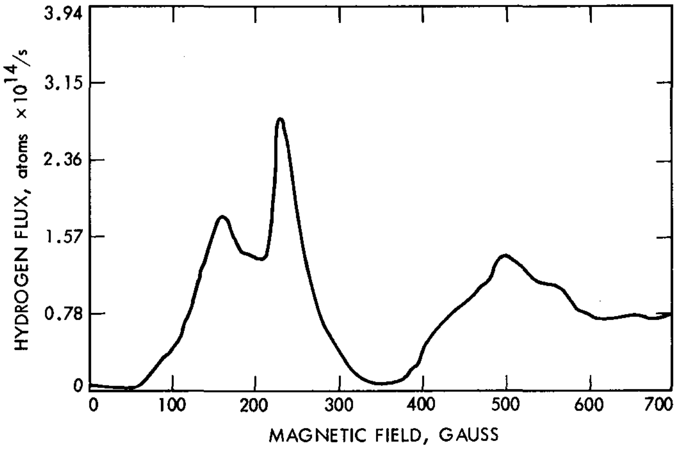


Fig. 2. Plot of flux of hydrogen atoms vs. magnetic field intensity as measured with the atom detector

X-Band Uplink Technology Demonstration At DSS-13

J. G. Meeker and C. T. Timpe
Telecommunications Systems Section

This article reports the status of the X-band uplink development program from a systems' viewpoint.

The hardware, designed and built under the program, is now in place at DSS-13. System stability testing is underway.

An X-band receive capability is being implemented on both the Galileo and the Venus Radar Mapper spacecraft. Experiments will be conducted that will serve both to demonstrate the capabilities of the X-band uplink system and to enhance science objectives on those missions. These objectives are briefly described.

I. Introduction

The DSN undertook a phased development program in 1979, under RTOP 64, to add an X-band uplink capability to the deep space network. The first phase was to design and build highly stable ground station subsystems for installation and checkout at DSS-13. A second phase was to arrange a flight demonstration of a complete X-band uplink system to verify predicted performance in the space environment.

In order to plan and coordinate the activities, a systems work unit was established. Under it, various tasks were accomplished, utilizing a design team representing the various hardware subsystems, space environment disciplines, and the user community. The main efforts were directed towards the following tasks:

(1) Quantifying the requirements imposed on the overall system by potential users of the X-band uplink, particularly in the navigation and radio science areas.

(2) Establishing error budgets for the telecom functions of telemetry, ranging and Doppler, including error estimates, with tolerances, for ground and spacecraft hardware, and the natural media effects due to the troposphere and interplanetary and ionospheric charged particles.

(3) Maintaining extensive liaison with NASA flight projects to develop mutually beneficial X-band uplink experiments.

(4) Preparing an Experiment Requirements Document, defining a flight demonstration of the X-band uplink, including preparatory ground testing and inflight calibrations as well as experiment sequences.

Much of the early systems work was summarized in TDA Progress Report 42-62 (Ref. 1).

Subsequently, the NASA spacecraft of the ISPM mission, which was initially to have carried the X-band uplink demonstration experiment, was cancelled. This report tells of the developments which have taken place since then, including the status of DSS-13 X-band uplink activities and plans of the DSN now to support X-band uplink operations on the Galileo and Venus Radar Mapper missions.

II. Status of X-Band Uplink Activities at DSS-13

A. Hardware Status

The ground station hardware (Ref. 2) has been built and tested at the subsystem level, and is now installed at DSS-13. Included are the following subsystems:

(1) The X-band uplink exciter, X/X and X/S Doppler extractors and mixers and test signal translators which coherently translate the X-uplink frequency to X- and S-downlink frequencies for in-station tests.

(2) The 20-kW klystron power amplifier, with buffer amplifier driver, complete with power supplies, controller and protective circuitry.

(3) A new microwave antenna feed design to accommodate S- and X-band uplinks and downlinks.

(4) An X-band diplexer, specially designed to minimize noise temperature contributions.

(5) Block III receiver modifications. In order to realize the potential advantages of the X-band uplink in reducing phase jitter due to charged particle variations in the solar wind and the ionosphere, the station will use a Hydrogen-maser timing reference and will incorporate a temperature-controlled Multi-mission Receiver (MMR) front end. Concepts consistent with unattended operations have also been employed.

The components that have been redesigned in the X-band uplink R & D effort, and are now in place at DSS-13, are indicated in the block diagrams. Fig. 1 shows the station as it would be configured to support an X-band demonstration with the Galileo mission, while Figs. 2 and 3 show the hardware in more functional detail for the X-band uplink/X-band downlink and X-band uplink/S-band downlink paths respectively.

B. DSS-13 System Test Status

Individual subsystems were checked out prior to installation at DSS-13 to verify that they met their design specifications.

In addition, a stepwise test plan was formulated to evaluate the total X-band uplink system. The approach emphasizes stability tests as the most informative. The testing is currently underway and will continue through 1984. The test sequence is as follows.

1. **Baseline exciter subsystem stability.** Exciter subsystem stability tests will serve as the baseline. Special tests were conducted prior to installation at DSS-13, in the Interim Frequency Standards Test Facility, using laboratory Hydrogen-masers as frequency references. Phase stability was measured at the X-band uplink frequency and at the downlink X- and S-band frequencies coherently derived from the uplink using the test signal translators that are part of the exciter subsystem. The phase data was processed to yield a square root Allan variance record versus sampling time.

The results of these measurements are encouraging. Fig. 4 shows a typical plot of the X-band uplink frequency stability over sampling times from 1 to over 20,000 seconds. The measurements have established that these exciter subsystem elements operate within the stability bounds specified for their design. Typical ($\Delta f/f$) values below 3 parts in 10^{-15} were obtained for integration times over 1000 seconds.

2. **X-band uplink stability at DSS-13.** The exciter stability measurements are being repeated at the station at this writing. Because of the hardware's location in the antenna feedcone, a portable, highly stable microprocessor data system had to be designed for these in situ measurements. This has been built and tested, verifying an acceptably close match in stability to the laboratory instrumentation described previously.

The uplink frequency stability measurements will next be extended to various points moving progressively from the exciter to the buffer amplifier and the 20-kW klystron power amplifier. These measurements will identify additional noise increments added along the path to the antenna.

The measurements will be performed first by open loop, and then by closed feedback path back to the exciter from, for example, the output of the klystron power amplifier. The feedback loop has been included to compensate for relatively slow phase changes in subsystems that are especially subject to variations in ambient temperature. The loop should essentially clamp the phase's excursions to those of the exciter, and its success depends on the stability in the feedback path and the gain in the loop. Note that the loop time constant should be long enough to permit passage of command and ranging modulations.

3. **Receiver testing.** The next tests will bring in portions of the receiver. For these tests, downlink frequencies will be generated using the X-up/X-down and X-up/S-down frequency translators operating on a sample of the X-band uplink extracted via a directional coupler. The receiver frequencies would be reinserted in the downlink path, also via couplers, after being adjusted to the proper level. Stability measurements would be made at these RF frequencies down to the point where downconversion occurs. At this point, viz the outputs of the X- and S-band masers, the DSS-13 station configuration is representative of the configuration which will be implemented in the High Efficiency (HEF) operational stations, DSS-15, DSS-45 and DSS-65. Therefore, stability measurements made at DSS-13 at the maser output should also be representative of those expected at the operational stations. From this point, however, the configurations differ. The HEF stations, shown in Fig. 5, will route closed-loop data for radiometric measurements, such as Doppler, through conventional 34-m station hardware consisting of an X- to S-downconverter followed by a Standard Block III S-band receiver.

In the HEF stations a separate signal path will utilize the temperature-controlled and phase-calibrated Multimission Receiver (MMR) RF frontend. This path leads to an open-loop receiver within an Occultation Data Assembly (ODA) for use by radio scientists to collect data for further processing in experiments such as the extraction of Doppler signatures for gravitational wave searches to be conducted on the Galileo mission.

In contrast, DSS-13 will utilize only a closed-loop receiver at this time, but will enhance the Block III receiver performance by adding modifications. The receiver modifications consist of a MMR frontend and an upconverter/downconverter section prior to interfacing with a Standard Block III receiver. Figs. 1 and 5 compare the configurations at DSS-13 and the HEF stations respectively.

4. **Doppler stability tests at DSS-13.** The next step is to bring the entire closed-loop DSS-13 system into play in an end-to-end stability test. The most meaningful point in the system to make this test is felt to be the output of the Doppler mixer, at either the 5 MHz or 1 MHz bias frequency point. This measurement would involve all of the transmit and receive hardware at the station, except the antenna, and would be indicative of the system noise "floor" for the station components.

5. **Inflight demonstration.** The final test of system stability will be accomplished using an actual space mission. This will first occur during the Galileo flight and will bring into play the elements of the total system not available before, including the station antenna, the total spacecraft receive/transmit system, and the intervening media.

Emphasis in the initial inflight demonstration test would again be on two-way Doppler stability using DSS-13 and Galileo. In order to obtain the best achievable conditions, the experiment would be performed at Galileo's first opposition. This view period has two advantages from a stability standpoint. It would provide a "quiet" environment (nighttime viewing) and would occur while the spacecraft is still close to earth; hence, it would feature strong signal conditions. The test should, in effect, establish a new system noise "floor," but this time for the total end-to-end system, complete with spacecraft and media.

Subsequent X-band uplink technology demonstration sequences would test other telecommunications functions, such as commanding and two-way coherent telemetry.

C. Supporting Analysis: Doppler Error Source Modeling

As noted, end-to-end tests are planned during 1984 to measure the phase stability of the DSS-13 X-band uplink station configuration. Preceding this, an analysis task was undertaken to guide the interpretation of the test results. In it, equations were derived for the phase transfer functions relating the contributions of error sources in the system to various output points including the X-band and S-band biased Doppler mixer outputs. A follow-up task is recommended for the future. This will involve obtaining and assigning phase power spectra for the major phase error sources such as VCOs, frequency synthesizers, cables and the H-Maser reference so that accurate quantitative predictions can be made.

The first part of this analysis, the derivation of the transfer-function equations, has been completed and was reported in TDA Progress Report 42-76 (Ref. 3). It considered two modes. One, for station testing, uses a test signal translator (a zero-delay device) as an uplink/downlink turnaround device. The second configuration models a system with an actual spacecraft at a significant round trip light time from the DSN station. The results indicate that system tests at the station can be useful in localizing many of the phase errors contributed by the station hardware, but cannot accurately reflect the net Doppler error that will occur during actual spacecraft tracking. Aside from space environment factors, lack of fidelity in simulating real mission conditions is due to two main differences between test and flight conditions:

(1) The round-trip times between an uplink transmission and subsequent downlink reception in a mission cannot be simulated in station tests. Therefore, in the station tests, phase error sources that are common to both the

uplink signal and the Doppler extractor will be highly correlated because they are generated from the same reference. These errors will be largely cancelled out in a process such as the conventional two-way Doppler measurement, wherein a Doppler reference and an estimate of a received signal are subtracted. However, some errors that enter only one part of the system, such as the phase error contributions of the X-band and S-band receiver VCOs and frequency synthesizers (DCOs) will be accurately identified in station tests.

(2) Translator turnaround devices do not simulate the spacecraft transponder function accurately. A transponder produces a downlink frequency by multiplying the uplink frequency by the required turnaround ratio (880/749 for X/X and 240/749 for X/S). On the other hand, the test signal translators, used in the DSS-13 implementation, pick off samples of the uplink frequency from selectable points in the exciter power amplifier chain, and, using mixers, simply add or subtract coherent components required to give the downlink X- or S-frequencies. These components of the downlink frequencies are derived directly from the H-maser and DCO references and are not subject to errors generated in the power amplifier section of the uplink.

Thus, an X- or S-downlink generated by a test signal translator would have essentially the phase noise of the uplink. This is not the same as found in a downlink generated by a transponder where the phase noise would be multiplicatively increased by the X-down/X-up turnaround ratio (880/749) or decreased by the S-down/X-up turnaround ratio (240/749). This error, of course, could be eliminated by using a transponder for station tests, when available.

The above conclusions are somewhat qualitative, but do point out what to expect. Within the restrictions noted, the phase transfer functions derived in the analysis can be used to calculate quantitative results, given models for the phase power spectrum of each of the phase error sources.

III. Flight Projects

Both the Galileo and Venus Radar Mapper (VRM) projects have made decisions to fly an X-band uplink receive capability. Support had been provided to them prior to, and since, that time by the X-band uplink system work unit. This consisted primarily of technical inputs, such as estimates of radiometric accuracies using X-band vs S-band uplinks, and preliminary estimates of spacecraft hardware impacts and costs.

The X-band uplink sequences during these two missions will serve two purposes. On the one hand they will compare the performance of telecom functions using an X-band uplink system to the predictions of the systems design team. And they are expected to enhance selected flight project science objectives. Because science enhancements, rather than operational objectives, are primary for the X-band uplink on these two missions, present plans call for single string (nonredundant) implementations of X-band receive hardware on both spacecraft.

Fortuitously, the science experiments on Galileo and VRM present the opportunity to test an X-band uplink system in *two different natural environments*, using different station configurations. The gravitational wave search by Galileo will be conducted during solar oppositions (nighttime viewing) at long range. Doppler signatures with slow variations are expected. Open-loop receivers and extensive processing will be used.

The VRM gravity experiment, on the other hand, will necessarily be conducted during daytime viewing, with the sun-earth-probe angle often quite small. The Doppler signature in this experiment will be acquired using the closed-loop receiver and the sampling time will be short, i.e., typically 5 seconds.

Some of the objectives achievable through the use of an X-band uplink are summarized:

A. Galileo: X-Band Uplink Demonstration and Gravitational Wave Search

The objectives to be accomplished with the use of X-band uplink on the Galileo Orbiter can be included in three general areas. They are:

(1) A demonstration of improved performance in the various telecommunication links relative to S-band uplink.

(2) A demonstration of the DSN capability to maintain flight operations with operational, high-power X-band transmitters.

(3) A significant enhancement in the capability to perform the Gravitational Wave Experiment.

The plan for the demonstration of the advantages of X-band over S-band in telecommunication link performance is essentially the same as that which was previously intended for the ISPM U.S. spacecraft. Since these objectives were defined in detail in a previous report, (Timpe, C. T., private communication) they will be summarized here. The telecommunications links on which improved performance will be demonstrated with X-band uplink are:

(1) Two-way coherent carrier tracking (Doppler stability).

(2) Command threshold.

(3) Two-way ranging.

(4) Two-way coherent telemetry data-rate.

The expected improvements in performance will result from the following two main factors:

(1) The net effect of a higher frequency on a one-way link (having the same size antennas, the same transmit power, and the same receive sensitivity) is a performance gain by a factor of approximately the square of the frequency ratio.

(2) The susceptibility of radiated energy to phase fluctuations induced by charged particles in the ionosphere, and in the interplanetary plasma decreases with higher frequencies.

The enhancement in the capability to perform the Gravitational Wave Experiment will result from the improvement in Doppler fractional frequency stability with X-band uplink. The system goal for overall fractional frequency stability on the ISPM ESA spacecraft with S-band uplink is 3×10^{-14} (the Allan variance on $\Delta f/f$ for a 1000 second time interval). The corresponding goal for Galileo with X-band uplink is 5×10^{-15}, or nearly an order of magnitude improvement. The two spacecraft will perform a joint gravitational wave (GW) search during their coincident solar opposition with the Earth in October of 1987. It is anticipated that GW activity, which may not be discernible in ISPM Doppler data by itself, may be very significant in verifying the existence of GW activity present in the Galileo Doppler data which can be time-correlated.

B. Venus Radar Mapper: X-Band Doppler Data for the VRM Gravity Experiment

In the past, S-band Doppler data have been the primary source for gravity field information. The experimenter relied strictly on what the navigation transponder produced. However, the major error in the S-band data is the effect from interplanetary media (plasma). The plasma produces amplitudes and frequencies that can be interpreted as gravity anomalies, when features near the resolution limits are being investigated.

With an X-band Doppler system these errors and high frequency noise can be reduced by a factor of 5 to 8 and much better resolution and statistical results on all features can be obtained. Presently, a gravity map of Venus has been extracted from the Pioneer Venus Orbiter data. A more detailed map will be obtained from Venus Radar Mapper data due to the X-band Doppler system and the smaller eccentricity of its orbit.

Gravity data are generally correlated with topographic results and then inferences are made on internal structure. For example, thickness of a crust or depth of isostatic adjustment are calculated at many areas. These in turn provide insight as to a general global structure. Also many topographic features can be classified by their gravity signatures, and, by using other geological evidence, inferences can be made on construction and age.

References

1. Komarek, T. A., Meeker, J. G., and Miller, R. L., ISPM X-Band Technology Demonstration. Part I, Overview, *TDA Progress Report 42-62*, Jet Propulsion Laboratory, Pasadena, California, January and February, 1981.

2. Hartop, R., Johns, C., and Kolbly R., X-Band Uplink Ground Systems Development, *Deep Space Network Progress Report 42-56*, Jet Propulsion Laboratory, Pasadena, California, January and February, 1980.

3. Koerner, M. A., Doppler Phase Transfer Functions for a System with an X-Band Uplink and X-Band and S-Band Downlinks, *TDA Progress Report 42-76*, Jet Propulsion Laboratory, Pasadena, California, October-December, 1983.

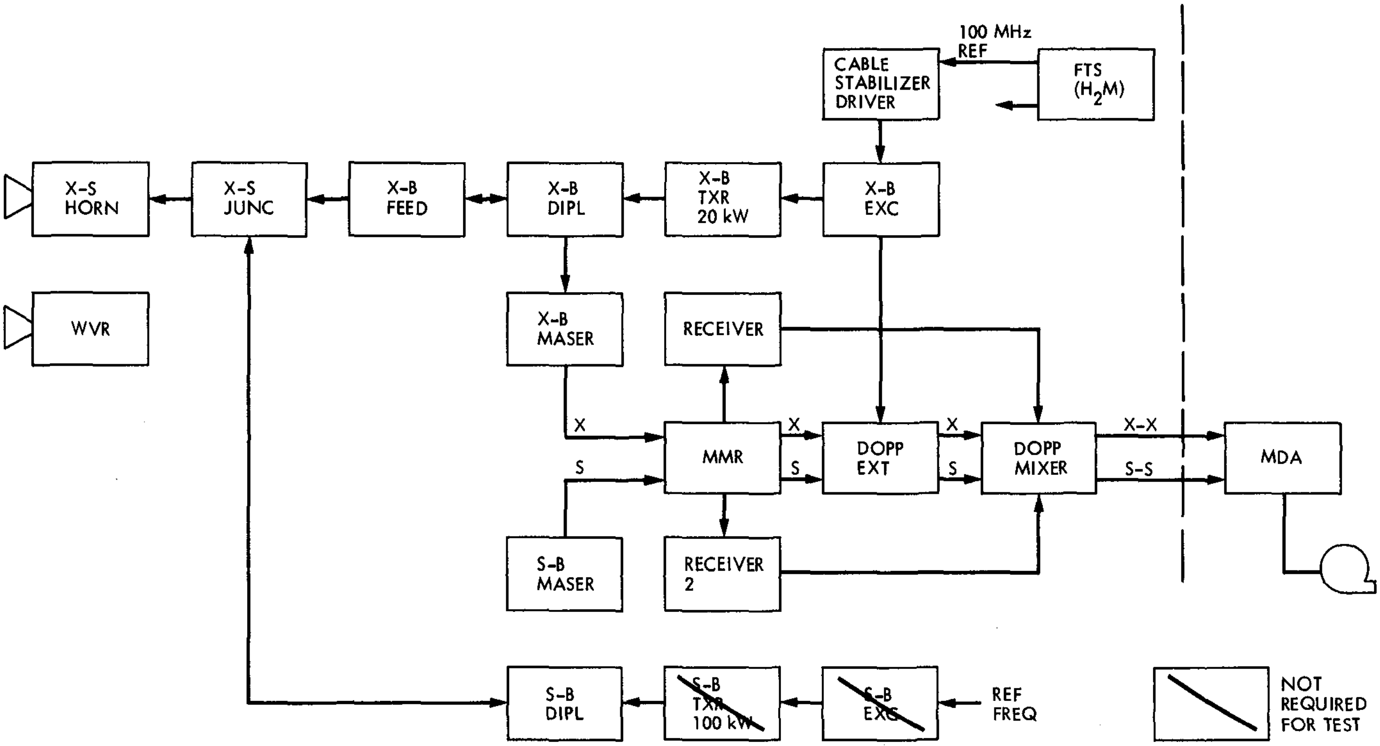


Fig. 1. X-band uplink functional configuration for DSS-13

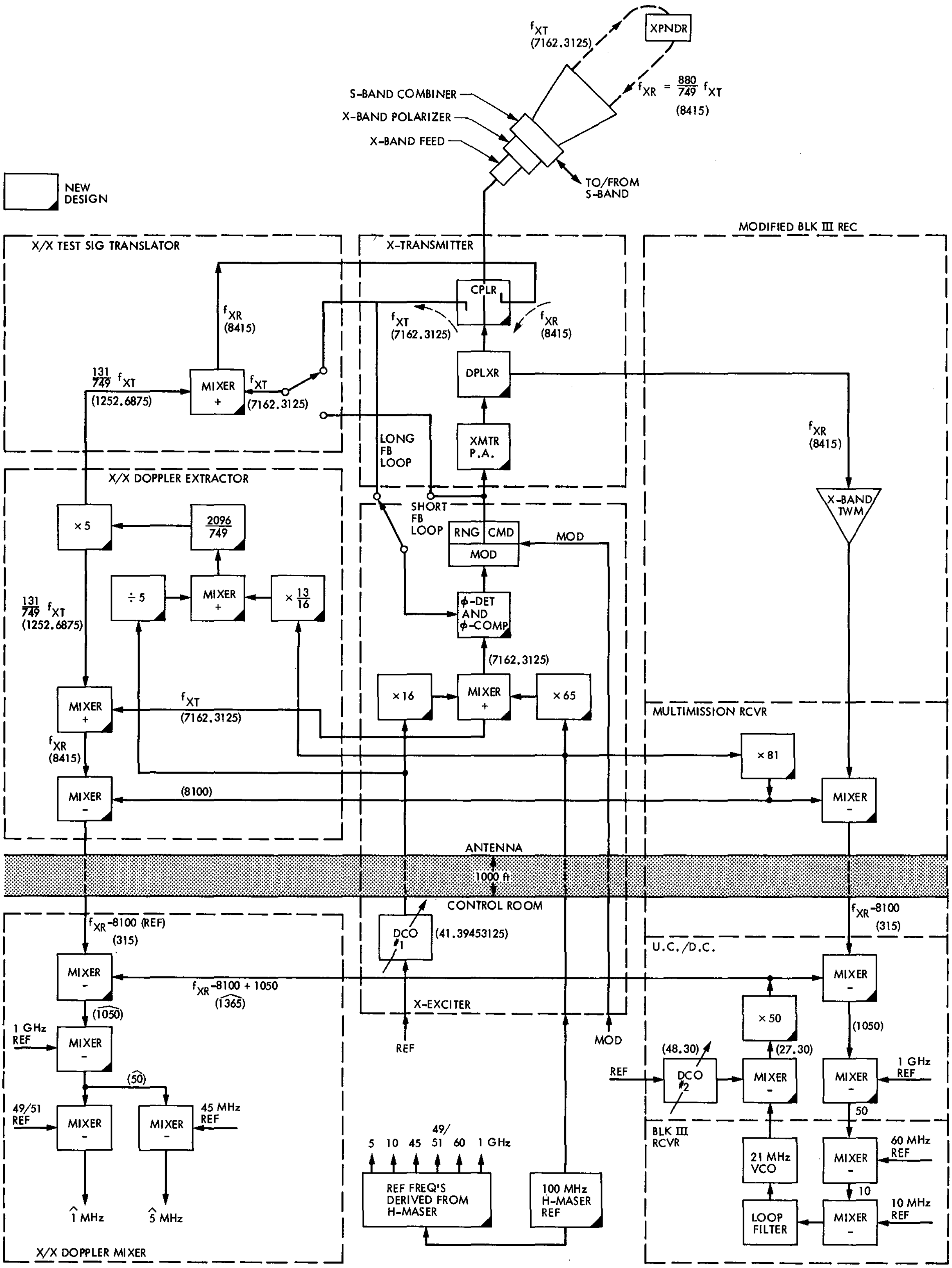


Fig. 2. Detailed functional block diagram for the X-band uplink/X-band downlink hardware at DSS-13 (Example frequencies are for channel 14)

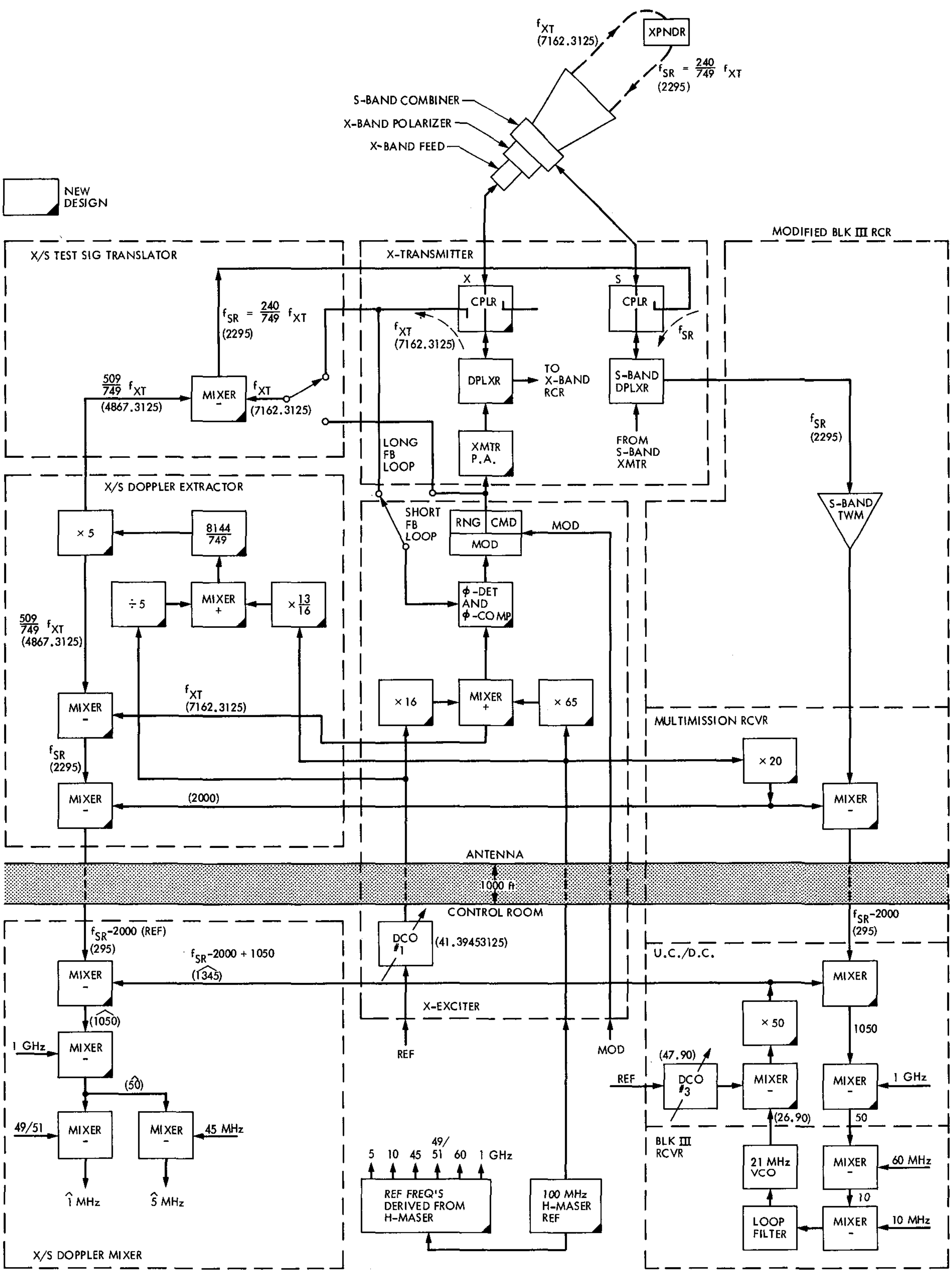


Fig. 3. Detailed functional block diagram for the X-band uplink/S-band downlink hardware at DSS 13 (Example frequencies are for channel 14)

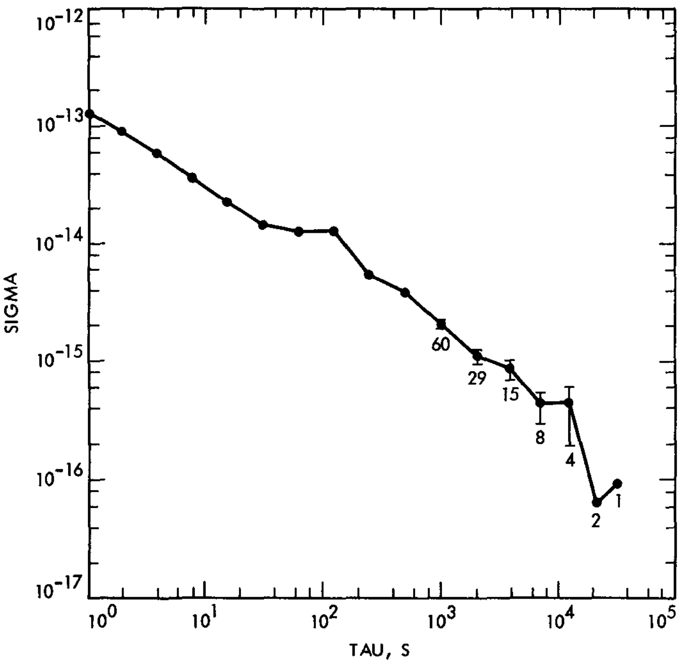


Fig. 4. X-band exciter fractional Allan variance at 7200 MHz, measured at the Interim Frequency Standards Test Facility

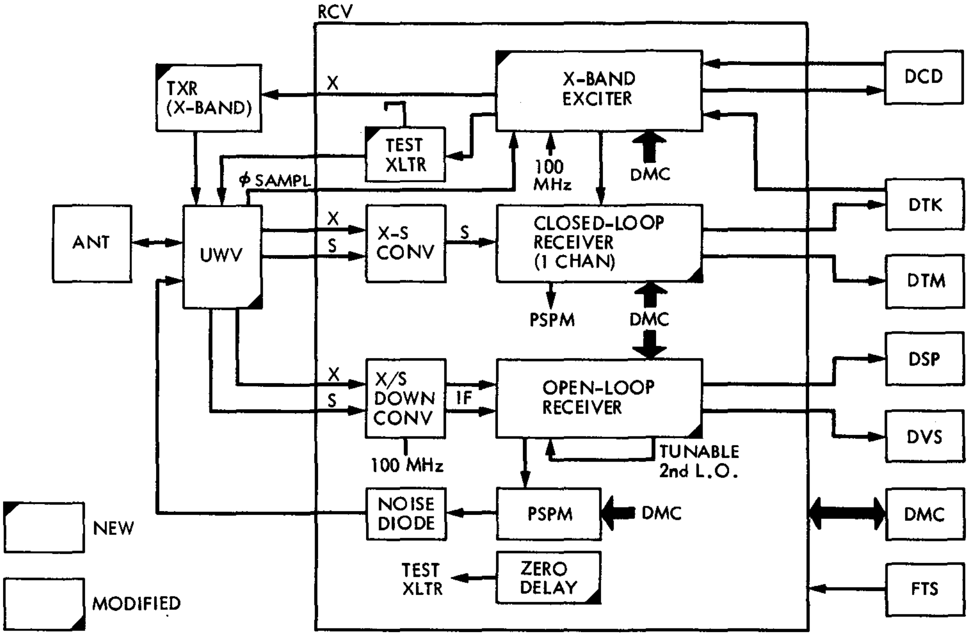


Fig. 5. Functional description block diagram for the 34-M HEF stations

Effects of NRZ-M Modulation on Convolutional Codes Performance

L. Deutsch, F. Pollara, and L. Swanson
Communications Systems Research Section

Non-Return-to-Zero Mark (NRZ-M) modulation is often used to resolve data sense in suppressed carrier telemetry systems. This is because such systems are subject to half cycle slips that result in complementing the encoded data stream. The performance of coded telemetry systems with NRZ-M is sensitive to the order in which the various operations are done. This means that a system that demodulates the NRZ-M waveform and then decodes will perform differently from a system that does the decoding first. In this report, the performance of the NASA standard (7, 1/2) convolutional codes is determined for several systems using NRZ-M. In addition, several different demodulation schemes for NRZ-M are considered. It is shown that, even for the best soft-decision method examined, there is a 2.7 dB loss at a decoded bit error rate of 5×10^{-3} if the NRZ-M demodulation occurs before rather than after Viterbi decoding.

I. Introduction

Emerging coding standards for interagency projects, such as those adopted by the Consultative Committee for Space Data Systems, will include Non-Return-to-Zero Mark (NRZ-M) modulation (Ref. 1) on suppressed carrier links. This type of modulation, used to avoid serious errors in case of half-cycle carrier synchronization slips, is defined as follows.

Suppose that $\{x_n\}$ is a binary stream of bits from $\{0,1\}$. This conventional binary representation of data is sometimes called "Non-Return-To-Zero Level" or NRZ-L modulation. The NRZ-M encoder output $\{y_n\}$ is defined by

$$y_n = x_n \oplus y_{n-1}$$

where $\oplus$ denotes addition modulo two.

Decoding (i.e., translation back to the original NRZ-L binary stream) is done by noting that $x_n = y_n \oplus y_{n-1}$.

The present DSN uses residual carrier transmission exclusively, which is subject only to full-cycle slips. In the DSN case NRZ-L modulation is therefore appropriate.

If a suppressed carrier system were used with NRZ-L modulation, a half-cycle slip, which corresponds to an inversion of the data stream, could easily destroy large blocks of data. With NRZ-M the inversion would cause only isolated errors. If there is an outer block code, such as the NASA standard Reed-Solomon code, then these isolated errors are likely to be corrected by the block decoder.

This article studies the effects of NRZ-M modulation on a (7, 1/2) convolutionally coded system. (The coding and decoding for such systems is not discussed here but a good presentation may be found in Ref. 2.) In particular, two different schemes for implementing NRZ-M will be examined. The first, called Inner NRZ-M, introduces NRZ-M encoder and decoder just before and after the channel, so that the NRZ-M modulation operates on convolutionally encoded bits. The second scheme introduces encoder and decoder before the convolutional encoder and after the Viterbi decoder, respectively, and will be called Outer NRZ-M. (Outer NRZ-M is an option only in the case of transparent codes [Ref. 2] such as the NASA standard convolutional codes.)

A detailed description of these two schemes is in Sections III and IV. Their performance with various decoding methods was determined by analysis and software simulation. These are described in Sections V and VI. The conclusions of the study follow immediately in Section II.

II. Conclusions

The addition of NRZ-M outside the convolutional channel (Fig. 1) causes only a slight drop in performance on a Gaussian channel whether symbols are hard- or soft-quantized, and so is a reasonable solution to the problem of half cycle slips in the case of suppressed carrier tracking.

The other possible implementation of NRZ-M, inside the convolutional channel (Fig. 2) causes far more performance degradation, and should be avoided.

Numerical results are shown and explained in Section VI.

III. NRZ-M Modulation

This type of modulation, widely used to avoid phase ambiguities of half cycles or multiples, is also and more properly referred to as "differential encoding" (Ref. 1).

Suppose that $\{x_n\}$ is the binary data stream to be encoded. Then the encoder output $\{y_n\}$ is defined as in Section I by:

$$y_n = x_n \oplus y_{n-1}$$

If we consider the sequence $\{x_n\}$ to be bipolar data (i.e., each x_n is taken from the set $\{-1,+1\}$ rather than $\{0,1\}$), then we may rewrite this expression as

$$y_n = -x_n y_{n-1}$$

This notation is used for the remainder of this article.

Let the received sequence be $\{\hat{y}_n\}$, where $\hat{y}_n$ is the result of passing y_n through a binary symmetric channel. In order to restore the original data, we must appropriately decode the received sequence $\{\hat{y}_n\}$. Let the NRZ-M decoder output be $\{\hat{x}_n\}$, where

$$\hat{x}_n = -\hat{y}_n \hat{y}_{n-1}$$

Then, if there are no errors in the received sequence, i.e., $\hat{y}_n = y_n$ for each n, we have

$$\begin{aligned}\hat{x}_n &= -\hat{y}_n \hat{y}_{n-1}\\ &= x_n y_{n-1} y_{n-1}\\ &= x_n\end{aligned}$$

as required. Note that if both y_{n-1} and y_n are complemented, we still have $\hat{x}_n = x_n$. This invariance under signal inversion is the property that makes NRZ-M an attractive algorithm for suppressed carrier systems.

The encoder/decoder algorithm just described can be directly used to implement the Outer NRZ-M scheme.

The same is true for the Inner NRZ-M, if the received data is hard quantized. Otherwise, for a soft quantized receiver, we need to modify the NRZ-M decoder in some way.

With NRZ-L data, alternate symbols from the convolutional encoder are usually inverted. This increases received symbol transition probability, resulting in improved symbol synchronization (Ref. 3). In the case of Outer NRZ-M encoding, such a system would increase symbol transitions in the same way, with no change in the results of this report. In the case of Inner NRZ-M encoding, alternate symbol inversion schemes do not guard against long strings without transitions.

IV. Soft-Decision NRZ-M Decoding

The implementation of Inner NRZ-M encoding/decoding requires either a new Viterbi decoder that operates on NRZ-M encoded symbols, or an algorithm for the NRZ-M decoder to send soft quantized symbols to the Viterbi decoder. Conventional NRZ-M decoding described in Section III produces hard-quantized symbols as an output, but use of these in the Viterbi decoder would cause a substantial loss in Viterbi decoder performance. An optimal system based on channel probabilities is computationally cumbersome.

One system which has been suggested is

$$\begin{aligned}\hat{x}_n &= -sgn(\hat{y}_n \hat{y}_{n-1})|\hat{y}_n|\\ &= -(\hat{y}_n \hat{y}_{n-1})/|\hat{y}_{n-1}| \qquad (1)\end{aligned}$$

This algorithm is the one used in the Symbol Synchronizer Assembly (SSA) in the DSN, according to the Symbol Synchronizer Assembly Technical Manual, Operations and Maintenance, and it agrees in sign with hard-quantized NRZ-M decoding.

Another system which agrees in sign with hard-quantized NRZ-M decoding is

$$\hat{x}_n = -\hat{y}_n \hat{y}_{n-1} \tag{2}$$

This second system, which appears more natural to us, performed better in the Viterbi decoder simulations (see Section VI).

V. A Theoretical Comparison of the Inner and Outer Schemes

In the case that received information over the channel is assumed to be hard quantized, a mathematical analysis of inner and outer NRZ-M systems at high SNRs is possible. This analysis consists of considering channel error sequences that produce Viterbi decoder bit errors. The probability of decoded bit error is then approximated by a sum over these sequences of their probability of occurrence multiplied by the number of bit errors caused by each one of them.

Consider the Inner NRZ-M scheme. Also, assume that the sequence that is sent is the all zero sequence. Since both Viterbi decoding and NRZ-M decoding are linear operations, there is no loss of generality in this assumption as long as the system is synchronized. Since we are only considering hard quantization in this analysis, the channel may be modeled as a binary symmetric channel (BSC) with transition probability p.

In the NRZ-L system, the Viterbi decoder (either JPL [Ref. 4] or Goddard conventions) is capable of correcting any error pattern of Hamming weight less than five (Ref. 4). A computer search was performed to determine sequences that cause bit errors when they are passed first through an NRZ-M decoder and then the Viterbi decoder. Since a weight two channel error pattern can only become at most a weight four pattern after NRZ-M decoding, all such patterns are corrected by the inner NRZ-M system. All weight three channel error patterns of length up to 16 symbols were checked by using a software Viterbi decoder. It was found that there are weight three sequences that produce bit errors in both the JPL and Goddard decoders if they were first NRZ-M decoded. In fact, there were sequences that produce eight and five bit errors, respectively, for these two codes. A complete listing of these sequences is shown in Table 1.

At very high SNRs, the probability of a weight three channel error sequence occurring at any time is approximately p^3. Suppose the weight three error sequences are $\{v_i\}$ $(i=1,2,\ldots,n)$ and that the number of decoded bit errors that each produces is $b(v_i)$. Then for very high SNRs, the probability of decoded bit error is given by

$$P_{\mathrm{bit}} = \sum_{i=1}^{n} b(v_i) p^3 (1-p)^{13}$$

The channel transition probability p is given by

$$\begin{aligned} p &= Q(\sqrt{2E_s/N_o}) \\ &= Q(\sqrt{E_b/N_o}) \end{aligned}$$

where E_s is the energy in a channel symbol and Q is the Gaussian error function

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-\nu^2/2) d\nu$$

For large x we may use the approximation (Ref. 5)

$$Q(x) = \frac{1}{\sqrt{2\pi}\, x} \exp(-x^2/2)$$

Let $\alpha = E_b/N_o$. Then, for large α we obtain

$$P_{\mathrm{bit}} = \left[\sum_{i=1}^{n} b(v_i)\right] (1/2\pi\alpha)^{3/2} \exp(-3\alpha/2)$$

As α becomes large, this probability behaves like its exponential term. Hence, asymptotically,

$$P_{\mathrm{bit}} = \exp(-3\alpha/2) \tag{3}$$

for both the JPL and Goddard codes. In comparison, these codes, without NRZ-M on the channel, behave asymptotically like

$$P_{\mathrm{bit}} = \exp(-5\alpha/2) \tag{4}$$

Suppose that the bit SNR required to achieve a certain error rate for the conventional system is β and that the SNR required to get the same performance in the Inner NRZ-M system is α. Then by Eqs. (3) and (4),

$$\exp(-3\alpha/2) = \exp(-5\beta/2)$$

or

$$\alpha = (5/3)\beta$$

This means that the inner NRZ-M scheme should behave about 2.2 dB worse than the NRZ-L system in the limit as SNR becomes infinite. For "reasonable" signal-to-noise ratios, the difference is even greater because of the terms other than the exponential terms. In particular, to obtain bit error probability 0.005, $\alpha = 2.04\beta$, about a 3 dB loss.

If NRZ-M decoding is added after the Viterbi decoder, then the bit errors produced by the decoder are at most doubled. This constant factor is not significant in asymptotic behavior. Therefore the Inner NRZ-M scheme should also behave 2.2 dB worse than Outer NRZ-M for very high SNRs.

VI. Simulation Results

In order to determine the performance of various telemetry system configurations involving NRZ-M and convolutional coding, software simulations were carried out. The software modules included a convolutional encoder, Viterbi decoder, NRZ-M encoder, NRZ-M decoders for each of the NRZ-M decoding algorithms described above, and a Gaussian channel simulator. These were written in C-Language and run on a VAX 11/750 computer under the UNIX operating system.

The configuration used for the simulation is shown in Fig. 3. The "generator" output is a stream of 0's and 1's. The "coder" is a convolutional encoder implementing one of the standard NASA (7,1/2) codes. The "display" routine compares the delayed information bits with the decoded bits and generates the statistics. In addition there are NRZ-M coders and decoders which may be inserted in various places.

Figure 4 shows all configurations used in the simulation. Results are shown in Figs. 5 and 6, with decoded bit error rate as a function of the bit signal-to-noise ratio, E_b/N_o, for each scheme. The simulations assume perfect carrier and subcarrier tracking and perfect Viterbi decoder node synchronization. The Viterbi decoder path memory length was 32 bits.

The probability of error P_E is measured on a random sequence of N = 1,000,000 bits or until the 95% confidence interval C, defined as

$$C = 1.96\sqrt{P_E(1-P_E)/N}$$

is less than 5% of P_E. Therefore the true P_E is inside the range $P_E \pm C$ with 95% confidence. Table 2 shows a comparison of the various schemes for bit error probabilities of 10^{-2}, 5×10^{-3}, and 10^{-3}.

As expected, the Outer NRZ-M scheme outperforms the Inner NRZ-M scheme, for both the JPL and Goddard codes, the latter being slightly better.

References

1. Lindsey, W. C., and Simon, M. K., *Telecommunication Systems Engineering*, Prentice-Hall Inc., 1973.
2. McEliece, R. J., *The Theory of Information and Coding*, Addison-Wesley, 1977.
3. Simon, M. K., and Smith, J. G., "Alternate Symbol Inversion for Improved Symbol Synchronization in Convolutionally Coded Systems," *IEEE Trans. Comm.*, Vol. COM-28, February, 1980, pp. 228-237.
4. Yuen, J. H., *Deep Space Telecommunications Systems Engineering*, Plenum, New York, 1983.
5. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*, Wiley, 1965.

Table 1. A list of Viterbi-decoded bit error sequences caused by the Inner NRZ-M system as a result of weight three channel symbol error sequences. Hard quantization is assumed.

JPL (7, 1/2) code		Goddard (7, 1/2) code	
Viterbi-Decoded Bit Error Sequence	Number of Errors	Viterbi-Decoded Bit Error Sequence	Number of Errors
1001000000000000	2	1001000000000000	2
1000000000000000	1	1000000000000000	1
1011000000000000	3	1100100000000000	3
1100000000000000	2	1100000000000000	2
1100100000000000	3	1011000000000000	3
1010100000000000	3	1110010000000000	4
1010000000000000	2	1111001000000000	5
1111001000000000	5	1110000000000000	3
1111110000000000	6	1111000000000000	4
1011011000000000	5	1101100000000000	4
1111000000000000	4	1010010000000000	3
1100111111000000	8	1010000000000000	2
		1010100100000000	4

Table 2. E_b/N_0 required to achieve a given P_E for different schemes

P_E	Goddard Code				JPL Code			
	NRZ-L Soft	Outer NRZ-M Soft	Inner NRZ-M Soft[a]	Inner NRZ-M Soft	NRZ-L Soft	Outer NRZ-M Soft	Inner NRZ-M Soft[a]	Inner NRZ-M Soft
10^{-1}	0.5	0.65	3.1	4.0	0.45	0.55	3.1	4.0
10^{-2}	1.8	1.9	4.55	5.65	1.85	1.95	4.55	5.6
5×10^{-3}	2.1	2.2	4.85	5.95	2.15	2.25	4.85	5.95

[a] Method of Eq. (2)

Fig. 1. Outer NRZ-M coding system

Fig. 2. Inner NRZ-M coding system

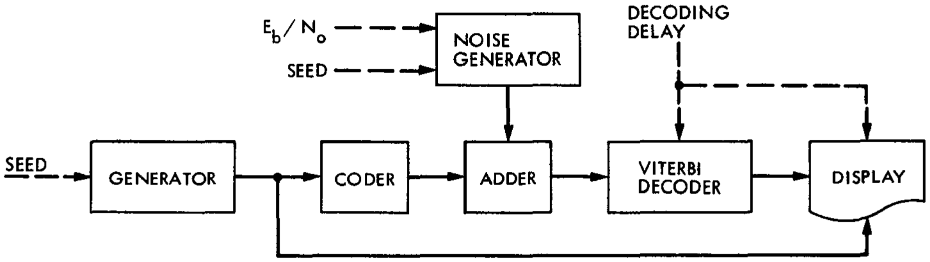


Fig. 3. Block diagram for simulation of NRZ-M with convolutional codes

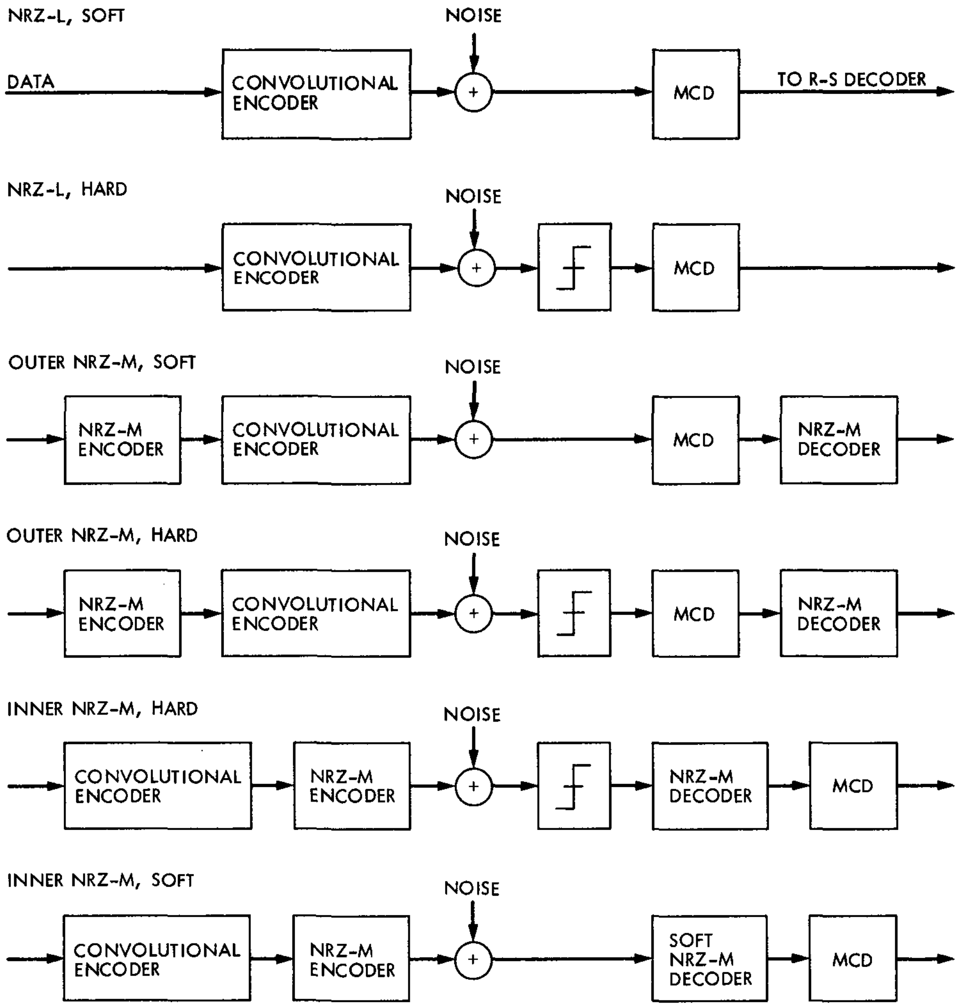


Fig. 4. Configurations used in simulation

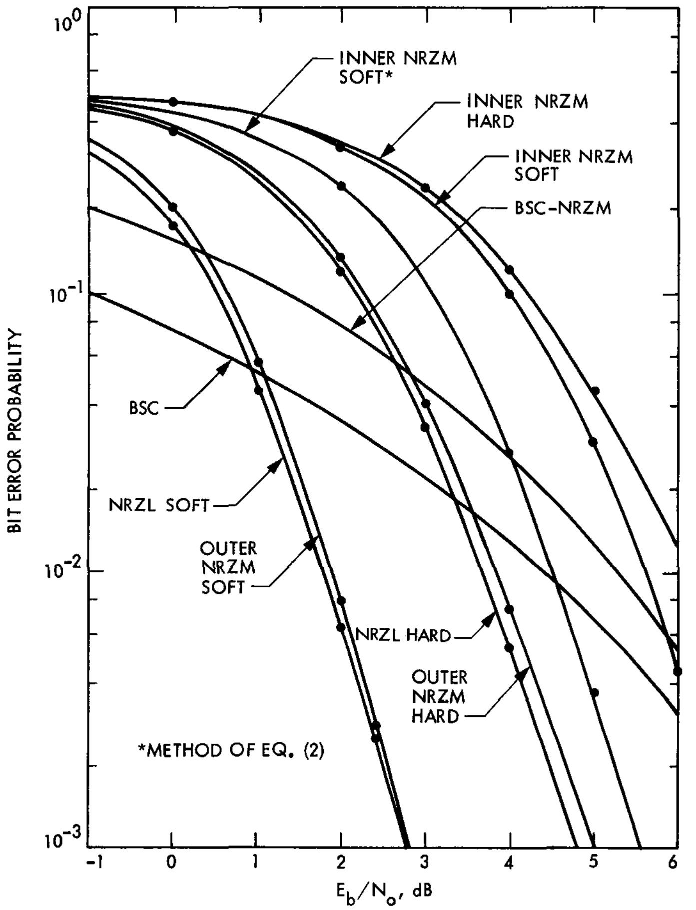


Fig. 5. Performance of Goddard (7, 1/2) convolutional code under various NRZ scenarios

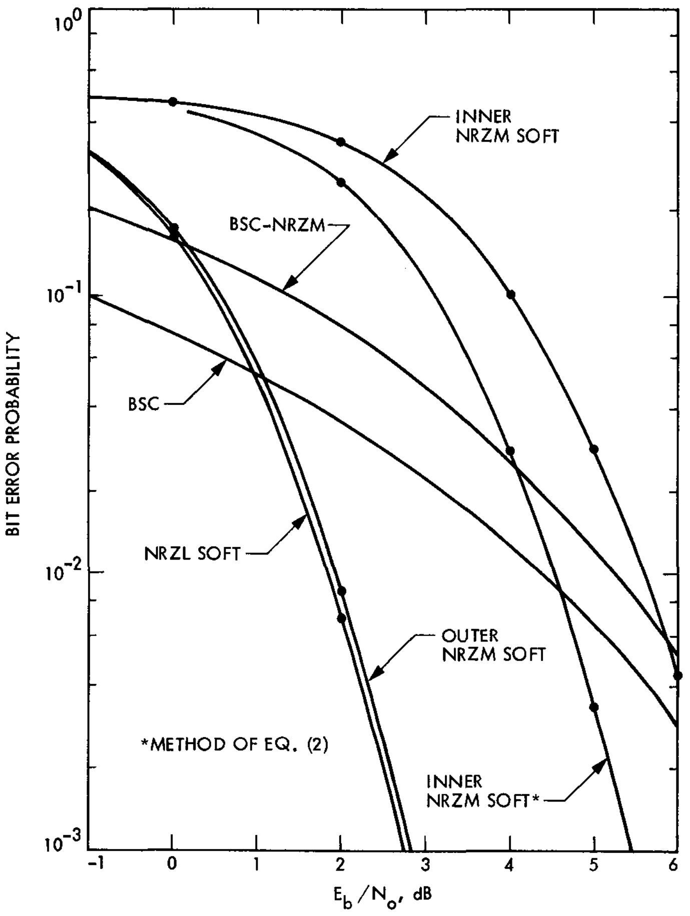


Fig. 6. Performance of JPL (7, 1/2) convolutional code under various NRZ scenarios

New Short Constraint Length, Rate 1/*N* Convolutional Codes Which Minimize Required E_b/N_o for Given Bit Error Rate

P. J. Lee
Communications Systems Research Section

Instead of using the criterion of maximum free distance, d_f*, or the maximum* d_f *with minimizing a few first distance profiles, we searched short constraint length rate 1/N convolutional codes using a new criterion of minimizing required bit energy-to-noise density ratio,* E_b/N_o*, for a given value of desired bit error rate (BER), for the goodness of a code. The considered channel was binary antipodal signaling over additive white Gaussian noise and no quantization at the channel output. For the BER calculations, the transfer function bounding technique was used. Partial searches were performed using some known facts and a very useful new idea that "good codes generate good codes." That is, for a given constraint length K, good rate 1/(N + 1) codes can be found by extending the code generator matrices of good rate 1/N codes. The code search results are tabulated for* $3 \leqslant K \leqslant 7$ *and* $2 \leqslant N \leqslant 8$*. For many pairs of K and N, the new codes are shown to save 0.1 to 0.4 dB in the required* E_b/N_o *compared to previously reported codes. Additionally, the benefits of coding bandwidth expansion are confirmed with our new codes.*

I. Introduction

The use of a short constraint length convolutional code together with Viterbi decoding has been very popular for several applications. The class of time-invariant, nonsystematic, rate 1/*N* convolutional codes has been studied much more extensively than any other class, partially due to its ease of analysis. The bit error rate (BER) at the Viterbi decoder output is well bounded by the well known transfer function bound (Refs. 1 and 2):

$$\text{BER} \leqslant c_o \cdot \frac{\partial}{\partial Z} T(D, Z) \bigg|_{D=D_o,\, Z=1} \tag{1}$$

where c_o and transfer function $T(D, Z)$ depend on the code and the type of channel used. D_o, which has a value between 0 and 1, is the union-Bhattacharyya distance of the coding channel (everything inside the encoder-decoder pair). The right hand side of Eq. (1) is often represented as a series expansion (Refs. 1 and 2) as

$$\mathrm{BER} \leqslant c_o \cdot \sum_{i=d_f}^{\infty} a_i \cdot D_o^i \qquad (2)$$

where d_f is the free distance of the code.

A large number of good rate $1/N$ convolutional codes have been found and reported (Refs. 3, 4, 5, etc.). Up to now, all the researchers have used the maximum d_f criterion or the criterion of maximum d_f together with minimizing the first few a_i's, for determining goodness of a code in their code search procedures. These criteria are very valuable, if the values of operating D_o's are much smaller than 1, since in such cases Eq. (2) can be well approximated by the first few terms. Hence, we may consider the previously reported codes to be good when the required BER is extremely small. However when the required BER is in the moderate range of 10^{-2} to 10^{-6}, such criteria may not be good since much more than the first few times are required for a good approximation to Eq. (2).

For uses with systems which require BER in the moderate range, we searched for good rate $1/N$ convolutional codes by using a new criterion of minimizing the bit energy to noise density ratio, E_b/N_o, required for a desired value of BER. For the evaluation of BER, we directly used the transfer function bound, Eq. (1). In all cases considered we assumed the use of binary antipodal signaling over the additive white Gaussian noise (AWGN) channel with no quantization at the channel output. For the desired values of BER, we picked 10^{-3} and 10^{-6}.

In the next section, the code structures and the transfer function bounding techniques are briefly reviewed for familization with our notation. Then the partial code searching techniques used in this study are presented, where the very important idea that "good codes generate good codes" is introduced and explained. The code search results are given for $3 \leqslant K \leqslant 7$ and $2 \leqslant N \leqslant 8$ and are compared with previously reported codes. As an example, for the $K = 7$ and $N = 4$ case, our new code can save 0.4 dB in required E_b/N_o compared to any previously reported code when the desired value of BER is 10^{-6}. We also confirm the benefits of coding bandwidth expansion with our new codes.

II. Preliminaries and Notations

A typical nonsystematic, constraint length K, rate $1/N$ convolutional encoder is shown in Fig. 1. The code connection box is often represented by an $N \times K$ binary matrix $\mathbf{G}$, which is called code generator matrix. For a given pair of K and N, this code generator matrix $\mathbf{G}$ determines the performance. Let

$$G(n) = (\mathbf{G}(n,1), \ldots, \mathbf{G}(n,k), \ldots, \mathbf{G}(n,K)), \quad n = 1, 2, \ldots, N \qquad (3)$$

Notice that in many reports (e.g., Refs. 4, 5), the code generator G is represented by $(G(i), \ldots, G(n), \ldots, G(N))$ and $G(n)$'s are in octal forms. We call this "a regular representation" of the code generator. For later use, we also define

$$g(k) = (\mathbf{G}(i,k), \ldots, \mathbf{G}(n,k), \ldots, \mathbf{G}(N,k)), \quad k = 1, 2, \ldots, K \qquad (4)$$

The nth bit in tth output vector y_n^t (see Fig. 1) for $n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots$ is given by

$$y_n^t = \sum_{k=1}^{K} \mathbf{G}(n,k) \cdot x^{t-k+1} \qquad (5)$$

where $\sum$ is a mod-2 summation, $x^t \in \{0,1\}$, $t = 1, 2, \ldots$ and $x^t = 0$ for $t < 1$ by convention. The "present state" at time t, S^t, is defined as

$$S^t = (x^{t-K+1}, \ldots, x^{t-1}) \qquad (6)$$

Notice that there are 2^{K-1} $(= M)$ distinct states for any value of N.

To find code transfer functions, one often uses state diagrams where nodes represent states and directed branches represent state transitions. The metric value on the directed branch from a state to another state (or to itself), assuming the existence of such a transition, is the product of D to the power of the Hamming weight of the output vector and Z raised to the Hamming weight of the input bit, when a binary input channel is used.

As an illustration, a $K = 3$, $N = 2$ convolutional encoder, with one of the best code connections, is depicted in Fig. 2. For this code,

$$G = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \quad G(1) = (1,1,1), \; G(2) = (1,0,1),$$

$$g(1) = (1,1), \quad g(2) = (1,0),$$

$$g(3) = (1,1)$$

and the regular representation of this code generator is (7,5). The number of states, M, is 4, and its state diagram is shown in Fig. 3.

The transfer function $T(D, Z)$ can be represented (Refs. 2 and 6) as

$$T(D, Z) = B \cdot (I - A)^{-1} \cdot C \qquad (7)$$

where I is the $(M - 1) \times (M - 1)$ unit matrix. The $(M - 1) \times (M - 1)$ matrix A, the $(M - 1)$-dimensional row vector B, and the $(M - 1)$-dimensional column vector C can be obtained from the state diagram as

$$\left.\begin{array}{rl} A(i, j) = & \text{metric value on the branch from state } i \\ & \text{to state } j \text{, if there is such a transition} \\ = & 0, \text{ otherwise} \\ B(j) = & \text{metric value on the branch from state 0} \\ & \text{to state } j \text{, if there is such a transition} \\ = & 0, \text{ otherwise} \\ C(i) = & \text{metric value on the branch from state } i \\ & \text{to state 0, if there is such a transition} \\ = & 0, \text{ otherwise} \end{array}\right\} \qquad (8)$$

where neither state i nor state j is the all zero state (state 0). For Eq. (1), we need (Refs. 2 and 6)

$$\frac{\partial}{\partial Z} T(D, Z) = \frac{\partial B}{\partial Z} \cdot (I - A)^{-1} \cdot C$$

$$+ B \cdot (I - A)^{-1} \cdot \frac{\partial A}{\partial Z} \cdot (I - A)^{-1} \cdot C \qquad (9)$$

For the previous example,

$$A = \begin{bmatrix} 0 & D & DZ \\ Z & 0 & 0 \\ 0 & D & DZ \end{bmatrix}, \quad B = [D^2 Z \ 0 \ 0], \quad C = \begin{bmatrix} 0 \\ D^2 \\ 0 \end{bmatrix}$$

$$(I - A)^{-1} = \frac{1}{1 - 2DZ} \begin{bmatrix} 1 - DZ & D & DZ \\ Z(1 - DZ) & 1 - DZ & DZ^2 \\ DZ & D & 1 - DZ \end{bmatrix}$$

$$T(D, Z) = \frac{D^5 Z}{(1 - 2DZ)}$$

and

$$\frac{\partial}{\partial Z} T(D, Z) = \frac{D^5}{(1 - 2DZ)^2}$$

The purpose of this report is to find good codes which minimize the required E_b/N_o for desired BER = 10^{-3} or 10^{-6}, assuming the use of binary antipodal signaling over an AWGN channel and no channel output quantization. For this special coding channel, D_o and c_o in Eq. (1) are given by (Refs. 1, 2)

$$D_o = \exp(- E_s/N_o) \qquad (10)$$

and

$$c_o = Q(2 \cdot d_f \cdot E_s/N_o) \exp(d_f \cdot E_s/N_o) \qquad (11)$$

where N_o is the one-sided noise power spectral density and E_s is the received signal energy per channel bit,

$$E_s = r \cdot E_b \qquad (12)$$

(= E_b/N, in our cases) and

$$Q(w) = \int_w^\infty \exp(- t^2/2) \cdot dt/\sqrt{2\pi} \qquad (13)$$

In the next section the techniques used for the partial code searches are described.

III. Code Searching Techniques

For rate $1/N$ convolutional codes of constraint length K, the number of possible code generators is $2^{K \cdot N}$. For example, for $K = 6$ and $N = 4$, the number is 2^{24} which is over ten million. Furthermore, because of our new criterion, to test the goodness of code generators, we have to perform matrix inversions which require considerable amounts of computing time. These make exhaustive searches prohibitively difficult except for cases of very small K and/or very small N. For moderate sizes of K and N, partial searches are necessary. In order to find good codes with partial searches, some techniques are required for reducing the code generator space effectively.

First, note that changes in the *orders* of $G(n)$'s will not change the state diagram at all (Fact 1) since the Hamming weights of the output vectors do not depend on the orders of their elements. When we use this fact, the required search space for code generators reduces by roughly a factor of N factorial. Also observe that *reversing the order* of the $g(k)$'s gives the same performance (Fact 2) since this is equivalent to just redefining the state in reverse order, i.e., $S^t = (x^{t-1}, \ldots, x^{t-K+1})$, instead of Eq. (6). This allows one to reduce the search space by roughly another factor of two. Notice that we can use these facts for reducing the search space with no loss in the chance of finding the best code. There are some other known facts on equivalence relations between code generators. However, in our partial code searches, only Facts 1 and 2 were used since the others are much more difficult to employ.

Recently, another useful observation was made in Ref. 7. That is, every known good convolutional code satisfies the following condition (in Ref. 7, it is called "flow conservation"); for each state in the state diagram, summation of the exponents of D in all incoming branches must be equal to sum of those in all outgoing branches. In Fig. 3, we can easily check that the code satisfies this condition. We noticed that this condition is automatically satisfied in our cases of rate $1/N$ if

$$g(i) = g(K) = (1, 1, \ldots, 1) \tag{14}$$

Notice that every reported good code in this class also satisfies Eq. (14). Hence we restricted our code searches to only those codes which satisfy Eq. (14). With this restriction, there may be some possibility of losing the optimum code. However, the use of Eq. (14) further reduces the required size of code search space by roughly a factor of 2^{2N}.

For a given pair of K and N (at least one of them is small), we established the code space to be searched using Eq. (14) and Facts 1 and 2. Then catastrophic codes were deleted, as were codes for which d_f was too small. We considered the value of d_f to be too small if

$$d_f < d_{fm} - \lceil (K \cdot N)/10 \rceil \tag{15}$$

where d_{fm} is the maximum free distance of $(K, 1/N)$ convolutional codes and $\lceil x \rceil$ is the smallest integer that is greater than or equal to x. For an example, consider the $K = 4$ and $N = 3$ case. The size of the reduced code space is 13 by using Eq. (14) and Facts 1 and 2, while the size of the original space is 4096 ($=2^{4 \cdot 3} = 2^{12}$). After deleting 5 catastrophic codes, the following are the code generators to be considered further, in regular representations with a natural ordering of the largest element to the left:

(17,17,15), (17,15,15), (17,15,13), (17,15,11),

(15,15,13), (15,15,11), (15,13,11), and (15,11,11)

In this case, d_{fm} is 10. Hence (15,11,11) is deleted since its d_f is 7.

For the pair of K and N, we estimated two values of E_b/N_o at which we believed the best code(s) could achieve the desired values of BER. Then for each of the remaining codes, we calculated the values of BER using Eq. (1) at those two values of E_b/N_o. Then we listed the codes with an ordering of the best one to the top. We consider a code to be better than another if the sum of the two common logarithmic values of calculated BER's is the smaller. Such a listing is illustrated in Table 1 for the above example of the $K = 4$ and $N = 3$ case.

By examining these listings for several values of N with a given (small) K, we made a very useful observation that *good codes generate good codes*. That is, for a given K, we can find good rate $1/(N + 1)$ codes by extending the code generator of good rate $1/N$ codes. For example, from the best (4,1/2) code, (17,15), we can obtain (17,17,15), (17,15,15), (17,15,13) and (17,15,11). Notice that all of these (4,1/3) codes are shown in Table 1. For more insight on this idea, still for $K = 4$, some of the upper parts of the listings for $N = 2$ to 5 are shown in Fig. 4, with lines indicating that the left codes (with smaller N) generate the right codes. We limited the number of listings in Fig. 4 so as not to complicate the figure. Since good codes are generated by good codes, we do not have to use all the rate $1/N$ codes for the generation of good rate $1/(N + 1)$ codes. For the example of $K = 4$, the 5 best rate 1/3 codes in Table 1 are generated by the 2 best rate 1/2 codes. Also, the 8 best rate 1/4 codes are obtained from the 4 best rate 1/3 codes, and so on. Hence when we use this idea, we can reduce the size of the code space very effectively.

This idea that "good codes generate good codes" can be supported by the following fact: If a code B is generated by a noncatastrophic code A with free distance d_{fA}, then code B is also noncatastrophic and its free distance, d_{fB}, is greater than d_{fA} by at least 2. The reason for this fact is as follows: The value of the exponent of D on a branch in the state diagram for code B is always greater than or equal to the value of the exponent of D on the same branch in the state diagram for code A. Hence the code B cannot be catastrophic, since the code A is noncatastrophic. Additionally, the values of the exponents of D on the departing branch from state 0 and on the incoming branch to state 0 for code B are greater than those for code A by one [see Eq. (14) and recall that the rate of code B is $1/(N + 1)$ while that of code A is $1/N$]. Therefore $d_{fB} - d_{fA}$ is always greater than or equal to 2.

Before presenting our code search results, we mention that we used the following approximation for the matrix inversion in Eq. (9);

$$(I - A)^{-1} = \sum_{m=0}^{\infty} A^m \approx \sum_{m=0}^{\nu} A^m \qquad (16)$$

and we picked the stopping number ν such that

$$\sum_i \sum_j A^m (i, j) < 10^{-5} \text{ for any } m \geqslant \nu \qquad (17)$$

Notice that the left hand side of Eq. (17) approaches 0 monotonically as m increases. With the approximation of Eqs. (16) and (17), we have 5 to 7 digits of accuracy for most values of E_b/N_o of interest.

IV. Results, Discussion, and Conclusions

Table 2 summarizes our code search results which give the best codes in the sense of minimizing the required E_b/N_o for desired BER = 10^{-6} and 10^{-3} (among the searched codes). Also, previously reported codes (in Refs. 3, 4 and 5) are compared. Note that some of the previously reported codes are shown to be the best with our criterion as well. In particular, the codes reported in Ref. 3 are very good since only $N = 2$ and 3 cases were considered and the criterion of maximum d_f with minimizing the first few a_i's in Eq. (2) was used. Note also that, as the code rate gets smaller (or N gets larger), the importance of maximum d_f diminishes, since the value of D_o gets larger for a given value of E_b/N_o [see Eqs. (10) and (12)]. An interesting point is that the (7,1/3) code reported in Ref. 3 happens to be the best for the desired BER = 10^{-6}, despite the fact that the larger d_f codes were accidently overlooked (as noted in Ref. 4). In Fig. 5, BER versus E_b/N_o plots are given for the $K = 7$ cases.

Using our new codes it is possible to save up to 0.4 dB in required E_b/N_o compared to previously reported codes [see the (7,1/4) case]. To see the gains visually, we plot the required E_b/N_o for BER = 10^{-3} and 10^{-6} versus N curves in Fig. 6. Note that for our new codes, the required E_b/N_o becomes smaller as the code rate gets smaller (Fact A) for a given (large) value of K. This fact A is known as the gain due to coding bandwidth expansion. That is, we can save in required signal energy at the expanse of using the larger bandwidth associated with a lower code rate (with similarly complex codes[1]). Note that for the previously found codes, this fact cannot be observed, since the codes are not necessarily optimum in the sense of minimizing the required E_b/N_o. Note also that for small values of K (3 and 4) this fact cannot be seen even with our codes. Therefore, we conjectured that the benefit of coding bandwidth expansion can be obtained only when enough complexity is allowed.

Now consider the capacity and cutoff rate for further insight into Fact A. The capacity, C, in the dimension of code rate [information bits/channel bit] of this coding channel (binary antipodal signaling over AWGN, no channel output quantization) is given by (Ref. 2)

$$C = [-(1 + \ln 2\pi)/2 - \int p(y) \cdot \ln p(y) \cdot dy] / \ln 2 \qquad (18)$$

where

$$p(y) = [\exp(-(y - b)^2/2) + \exp(-(y + b)^2/2)] / \sqrt{8\pi} \qquad (19)$$

and

$$b^2/2 = E_s/N_o \qquad (20)$$

The channel coding theorem (Ref. 2) says, with coding, arbitrary small error rate can be achieved provided that the data rate R_b [information bits/sec] is less than the channel capacity $\mathscr{C}$ [information bits/sec] of the coding channel. With dimensions of code rate, the above condition can be restated as

$$r < C \qquad (21)$$

That is, the choice of code rate must be smaller than C, in theory. Note that C is a function of E_s/N_o, and $C(E_s/N_o)$ is a monotonically increasing function of its argument. Hence its inverse function exists. Therefore Eq. (21) can be rewritten as

$$E_s/N_o > C^{-1}(r) \qquad (22)$$

With Eq. (12),

$$E_b/N_o > C^{-1}(r)/r \qquad (23)$$

We call the right hand side of Eq. (23) "the required E_b/N_o to achieve capacity." These values were found and plotted in Fig. 7 as a function of coding bandwidth expansion factor (or 1/code rate r). Also the cutoff rate of the coding channel R_o [information bits/channel bit] is given by (Ref. 2)

$$R_o = 1 - \log_2 (1 + \exp(-E_s/N_o)) \qquad (24)$$

---

[1] The complexity of $(K, 1/N)$ convolutional code increases as K increases and also as N increases. But since K is a much more important factor for complexity than N, we will consider the codes with the same K to have similar complexities for any N.

It is often said (e.g., Ref. 8) that a small enough error rate can be achieved, *with practical coding*, provided that R_b is less than the cutoff rate $\mathscr{R}_o$ in [information bits/sec]. Or, equivalently,

$$r < R_o(E_s/N_o) \tag{25}$$

This can likewise be inverted to yield

$$E_b/N_o > R_o^{-1}(r)/r = -\frac{1}{r} \cdot \ln\left[2^{(1-r)} - 1\right] \tag{26}$$

We call the right hand side of Eq. (26) "the required E_b/N_o to achieve cutoff rate." These values are also plotted in Fig. 7. These theoretical curves tell us that to be in the region below the capacity curve is impossible in theory; to be in the region between the capacity and the cutoff rate curves is possible in theory but difficult in practice; to be in the region above the cutoff rate curve is practically possible. Note that from both the capacity and the cutoff rate curves, we see that lowering the code rate gives the benefit of reducing the required E_b/N_o, which is the theoretical view of Fact A.

In the same figure, we have plotted the performances of our new codes for $K = 5, 6$ and 7 (from Fig. 6). From comparisons of the slopes of the curves for BER = 10^{-3} with those for BER = 10^{-6} cases, we conclude that the benefit of coding bandwidth expansion is greater when the desired BER is larger. Also, increasing the system complexity (i.e., increasing K) provides more gain when the desired BER is smaller, as seen by the larger spacing between the curves for the BER = 10^{-6} cases compared to the BER = 10^{-3} cases. Note also the similarity between curves for our new codes with those derived from theory (especially the cutoff rate curve).

In conclusion, we have found good $(K, 1/N)$ codes which minimize the required E_b/N_o for the desired BER = 10^{-3} and 10^{-6} for $3 \leqslant K \leqslant 7$ and $2 \leqslant N \leqslant 8$. For the partial searches of codes, we used some known facts together with another very useful idea that "good codes generate good codes." For many pairs of K and N, our new codes are shown to save 0.1 to 0.4 dB in required E_b/N_o for moderate values of required BER, compares to the previously reported codes. Also, we confirmed the benefits of coding bandwidth expansion with our new codes, whereas the previously reported codes did not uniformly confirm this property.

References

1. A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems," *IEEE Trans. Comm. Tech.*, Vol. COM-19, pp. 751-772, October 1971.
2. A. J. Viterbi and J. K. Omura, *Principles of Digital Communications and Coding*, McGraw-Hill Book Co., New York, NY, 1979.
3. J. P. Odenwalder, *Optimal Decoding of Convolutional Codes*, Ph.D. Thesis, Dept. Sys. Sci., Sch Eng. Appl. Sci., Univ. Calif., L.A., Los Angeles, CA, 1970.
4. K. L. Larsen, "Short Convolutional Codes with Maximal Free Distance for Rates 1/2, 1/3, and 1/4," *IEEE Trans. Info. Theory*, Vol. IT-19, pp. 371-372, May 1973.
5. D. G. Daut, J. W. Modestino, and L. D. Wismer, "New Short Constraint Length Convolutional Code Constructions for Selected Rational Rates," *IEEE Trans. Info. Theory*, Vol. IT-28, pp. 794-800, Sept. 1982.
6. J. K. Omura and M. K. Simon, *Modulation/Demodulation Techniques for Satellite Communications, Part IV: Appendices*, JPL Publication 81-73, Jet Propulsion Lab., Pasadena, CA, Nov. 1981.
7. R. Polazzo, *Analysis of Periodic Nonlinear Trellis Codes*, Ph.D. Thesis, Dept. Sys. Sci., Sch. Eng. Appl. Sci., Univ. Calif., L.A., Los Angeles, CA, 1984.
8. J. L. Massey, "Coding and Modulation in Digital Communications," *Proc. Intern. Zurich Seminar on Digital Comm.*, Zurich, Switzerland, March 1974.

Table 1. Examples of listing of code search results for the $K = 4$ and $N = 3$ case

G in octal	d_f	$\log_{10}$ (BER) at E_b/N_o = 6.0 dB	$\log_{10}$ (BER) at E_b/N_o = 3.5 dB	Sum of the Two Values
(17, 15, 13)	10	-6.059	-3.070	-9.129
(17, 15, 11)	9	-6.008	-3.082	-9.090
(15, 13, 11)	8	-5.609	-3.014	-8.623
(17, 15, 15)	10	-5.702	-2.567	-8.289
(15, 15, 13)	9	-5.516	-2.510	-8.026
(15, 15, 11)	8	-5.422	-2.490	-7.912
(17, 17, 15)	8	-5.279	-2.488	-7.767

Table 2. Code search results

K	N	Code Generator G in Regular Representation, octal	d_f	Required E_b/N_o, dB for desired BER=		Notes
				1.E-6	1.E-3	
3	2	7, 5	5	6.706	4.017	1, a, b
3	3	7, 7, 5	8	6.736	4.140	1, a, b
3	4	7, 7, 7, 5	10	6.938	4.354	2
3	4	7, 7, 5, 5	10	6.706	4.017	4, a, b
3	5	7, 7, 7, 5, 5	13	6.669	4.055	3, a, b
3	6	7, 7, 7, 7, 5, 5	16	6.736	4.140	3
3	6	7, 7, 7, 5, 5, 5	15	6.706	4.017	4, a, b
3	7	7, 7, 7, 7, 5, 5, 5	18	6.664	4.035	3, a, b
3	8	7, 7, 7, 7, 7, 5, 5, 5	21	6.685	4.080	3, a
3	8	7, 7, 7, 7, 5, 5, 5, 5	20	6.706	4.017	4, b
4	2	17, 15	6	6.180	3.735	1, a, b
4	3	17, 15, 13	10	5.958	3.437	1, a
4	3	17, 15, 11	9	5.994	3.427	4, b
4	4	17, 15, 15, 13	13	6.004	3.511	2
4	4	17, 15, 13, 11	12	5.906	3.286	4, a, b
4	5	17, 17, 15, 15, 13	16	5.991	3.483	3
4	5	17, 15, 15, 13, 11	15	5.909	3.332	4, a, b
4	6	17, 17, 15, 15, 13, 13	20	5.958	3.437	3
4	6	17, 17, 15, 15, 13, 11	19	5.865	3.325	4, a, b
4	7	17, 17, 15, 15, 15, 13, 13	23	5.974	3.462	3
4	7	17, 17, 15, 15, 13, 13, 11	22	5.849	3.306	4, a, b
4	8	17, 17, 17, 15, 15, 15, 13, 13	26	5.972	3.456	3
4	8	17, 17, 15, 15, 15, 13, 13, 11	25	5.860	3.332	4, a
4	8	17, 17, 15, 15, 13, 13, 11, 11	24	5.906	3.286	4, b
5	2	35, 23	7	5.745	3.495	1, a
5	2	31, 23	6	5.845	3.430	4, b
5	3	37, 33, 25	12	5.395	3.115	1, a, b
5	4	37, 35, 33, 25	16	5.303	2.999	2
5	4	37, 35, 25, 23	15	5.298	3.000	4, a
5	4	35, 31, 27, 23	14	5.317	2.965	4, b
5	5	37, 35, 33, 27, 25	20	5.270	2.923	3
5	5	37, 35, 31, 27, 25	19	5.243	2.924	4, a
5	5	37, 35, 33, 25, 21	18	5.297	2.912	4, b
5	6	37, 35, 35, 33, 27, 25	24	5.291	2.957	3
5	6	37, 35, 33, 27, 25, 23	23	5.211	2.880	4, a
5	6	37, 35, 33, 27, 25, 21	22	5.294	2.867	4, b
5	7	37, 35, 35, 33, 27, 27, 25	28	5.286	2.955	3
5	7	37, 35, 33, 31, 27, 25, 23	26	5.211	2.845	4, a
5	7	37, 35, 33, 27, 25, 23, 21	25	5.256	2.839	4, b
5	8	37, 37, 35, 33, 33, 27, 25, 25	32	5.284	2.949	3
5	8	37, 35, 35, 33, 31, 27, 25, 23	30	5.211	2.860	4, a
5	8	37, 35, 33, 31, 27, 25, 23, 21	28	5.280	2.819	4, b

Table 2 (contd)

K	*N*	Code Generator **G** in Regular Representation, octal	d_f	Required E_b/N_o, dB for desired BER=		Notes
				1.E–6	1.E–3	
6	2	75, 53	8	5.310	3.289	1
6	2	77, 45	8	5.236	3.242	4, a
6	2	75, 57	8	5.293	3.211	4, b
6	3	75, 53, 47	13	4.918	2.900	1, a
6	3	75, 67, 41	12	5.034	2.854	4, b
6	4	75, 71, 67, 53	18	4.836	2.747	2
6	4	77, 73, 55, 45	18	4.779	2.729	4, a
6	4	77, 73, 51, 45	17	4.807	2.719	4, b
6	5	75, 73, 71, 65, 57	22	4.826	2.680	3
6	5	77, 73, 71, 55, 45	22	4.742	2.660	4, a
6	5	75, 71, 65, 57, 53	22	4.753	2.645	4, b
6	6	75, 73, 65, 57, 55, 47	27	4.764	2.616	3
6	6	77, 73, 67, 55, 51, 45	26	4.694	2.591	4, a
6	6	75, 71, 65, 57, 53, 47	26	4.704	2.590	4, b
6	7	75, 75, 67, 65, 57, 53, 47	32	4.762	2.630	3
6	7	77, 73, 67, 63, 55, 51, 45	30	4.696	2.565	4, a
6	7	75, 71, 65, 57, 53, 47, 43	29	4.717	2.564	4, b
6	8	75, 73, 67, 65, 57, 57, 51, 47	36	4.728	2.599	3
6	8	77, 73, 67, 63, 57, 55, 51, 45	35	4.693	2.552	4, a
6	8	77, 73, 67, 63, 55, 51, 45, 41	32	4.772	2.541	4, b
7	2	171, 133	10	4.802	3.036	1, a
7	2	161, 133	9	4.818	3.035	4, b
7	3	175, 145, 133	15	4.599	2.706	2
7	3	171, 145, 133	14	4.489	2.672	1, a
7	3	161, 135, 107	13	4.600	2.666	4, b
7	4	163, 147, 135, 135	20	4.761	2.814	2
7	4	175, 151, 133, 117	20	4.372	2.520	4, a
7	4	173, 167, 135, 111	20	4.433	2.511	4, b
7	5	175, 147, 135, 135, 131	25	4.562	2.653	3
7	5	175, 151, 133, 127, 117	25	4.310	2.444	4, a
7	5	175, 165, 151, 133, 117	25	4.350	2.438	4, b
7	6	173, 163, 151, 137, 135, 135	30	4.394	2.499	3
7	6	175, 171, 151, 133, 127, 117	30	4.286	2.387	4, a
7	6	175, 165, 151, 137, 133, 117	30	4.307	2.385	4, b
7	7	173, 165, 147, 145, 137, 135, 135	36	4.381	2.456	3
7	7	175, 171, 155, 127, 123, 117, 113	34	4.266	2.356	4, a
7	7	175, 171, 165, 151, 133, 127, 117	35	4.282	2.353	4, b
7	8	173, 165, 153, 147, 137, 135, 135, 111	40	4.312	2.387	3
7	8	175, 171, 165, 151, 133, 127, 117, 113	39	4.251	2.327	4, a, b

Notes:

1: Found by Odenwalder (Ref. 3)
2: Found by Larsen (Ref. 4)
3: Found by Daut, et al. (Ref 5)
4: Found by the author
a: Minimizes required E_b/N_o for desired BER = 1.E–6
b: Minimizes required E_b/N_o for desired BER = 1.E–3

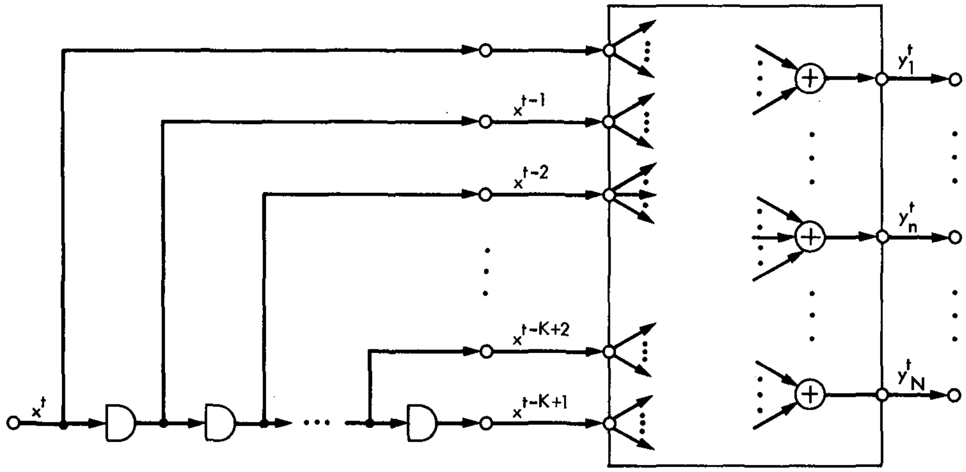


Fig. 1. A nonsystematic, constraint length *K*, rate 1/*N* convolutional encoder structure

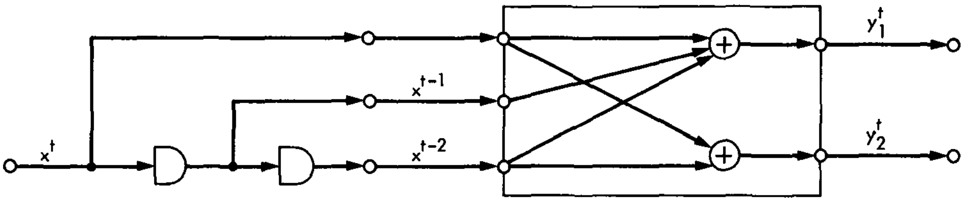


Fig. 2. A best (3,1/2) convolutional encoder

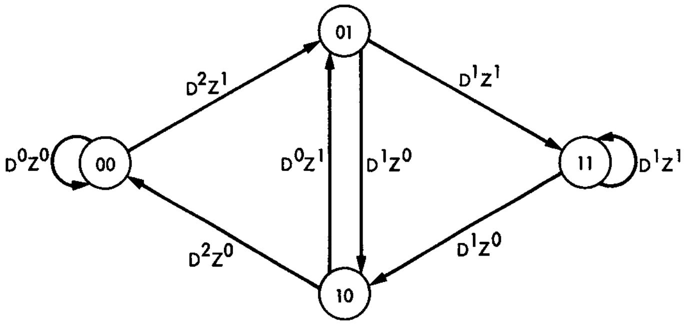


Fig. 3. The state diagram of the encoder in Fig. 2

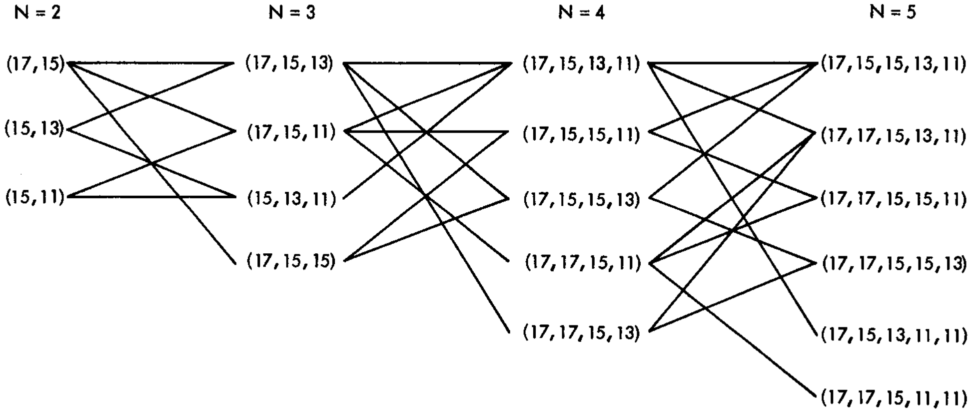


Fig. 4. An example for the idea of "good codes generate good codes," $K = 4$

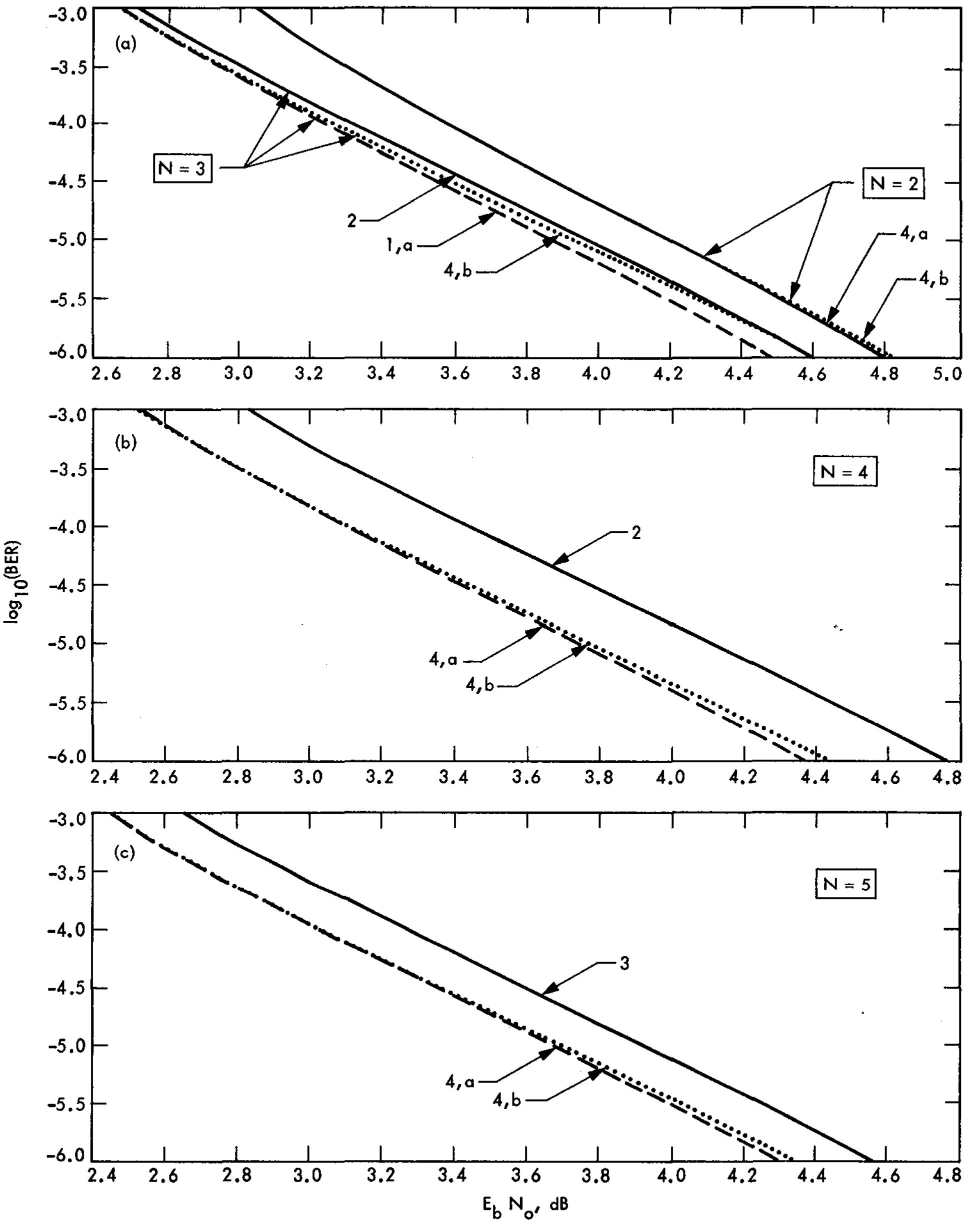


Fig. 5. BER versus E_b/N_0 curves for $K = 7$ codes (see Table 2 for the notes)

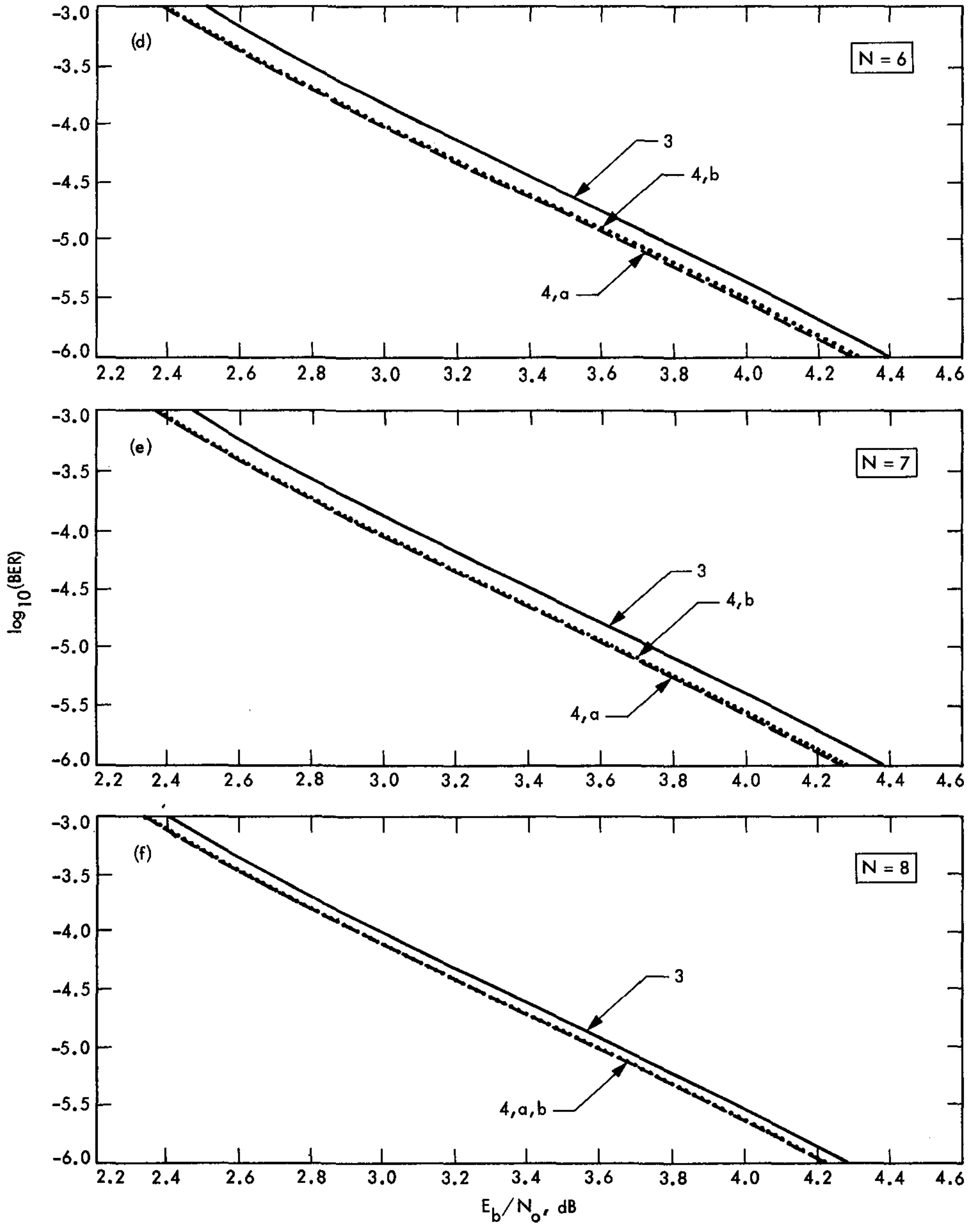
(d)
N = 6
3
4,b
4,a
(e)
N = 7
3
4,b
4,a
(f)
N = 8
3
4,a,b
-3.0
-3.5
-4.0
-4.5
-5.0
-5.5
-6.0
$\log_{10}$(BER)
2.2
2.4
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
E_b/N_o, dB

Fig. 5 (contd)

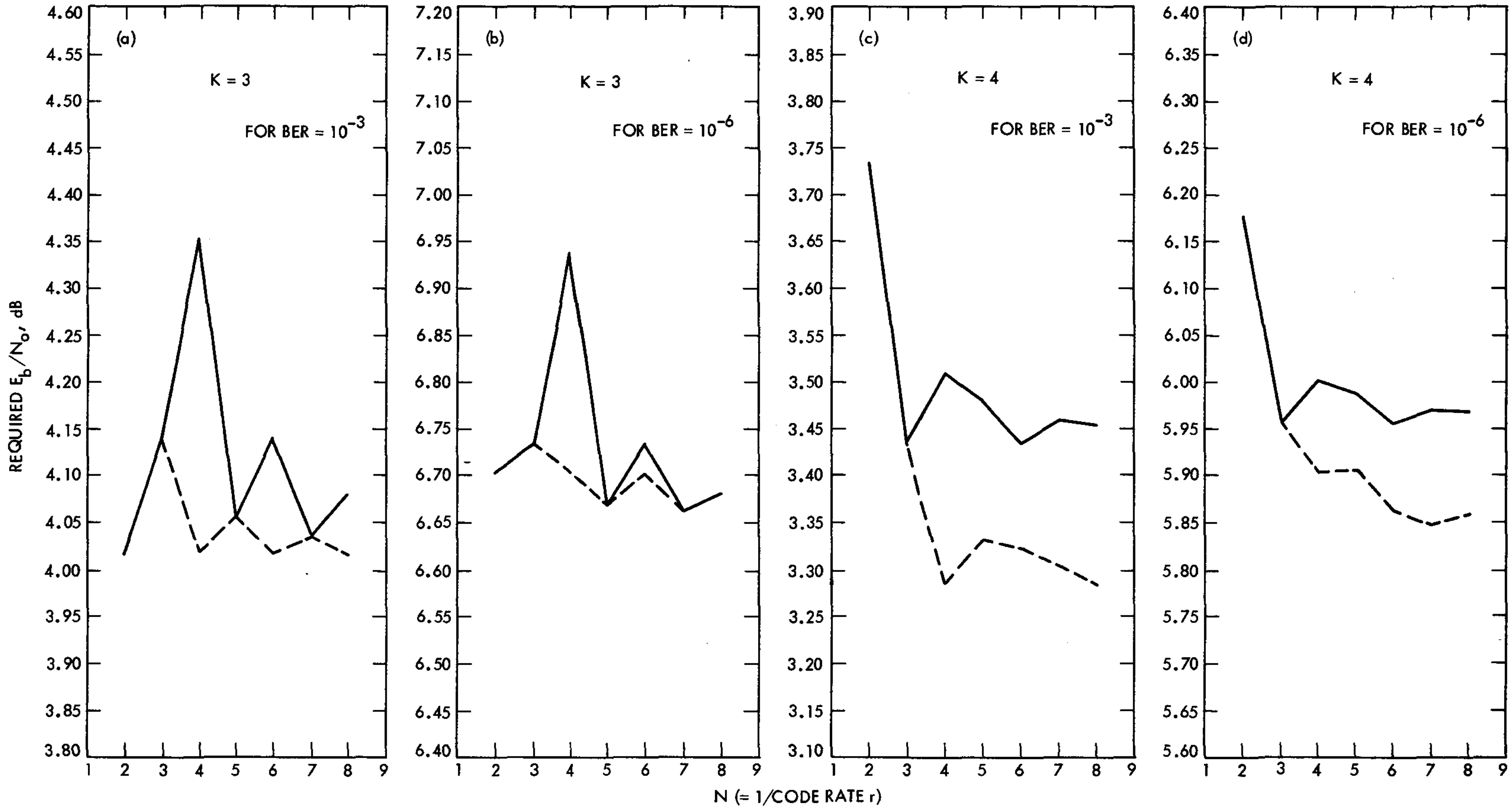


Fig. 6. Required E_b/N_o versus N (= 1/code rate r). Solid line indicates codes found previously; dashed line indicates codes found here.

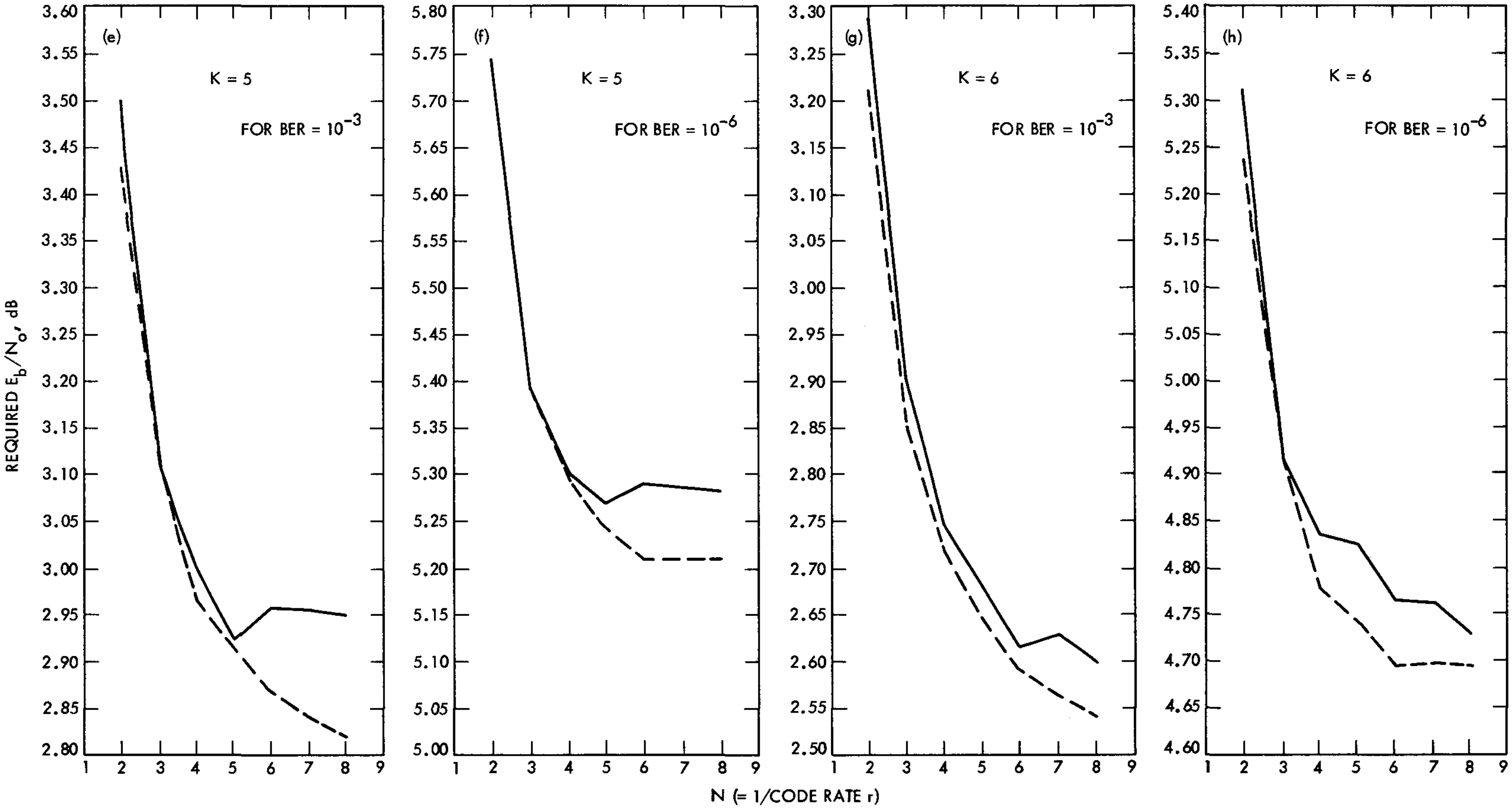

REQUIRED E_b/N_o, dB
(e)
K = 5
FOR BER = 10^{-3}
3.60
3.55
3.50
3.45
3.40
3.35
3.30
3.25
3.20
3.15
3.10
3.05
3.00
2.95
2.90
2.85
2.80
(f)
K = 5
FOR BER = 10^{-6}
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
5.35
5.30
5.25
5.20
5.15
5.10
5.05
5.00
(g)
K = 6
FOR BER = 10^{-3}
3.30
3.25
3.20
3.15
3.10
3.05
3.00
2.95
2.90
2.85
2.80
2.75
2.70
2.65
2.60
2.55
2.50
(h)
K = 6
FOR BER = 10^{-6}
5.40
5.35
5.30
5.25
5.20
5.15
5.10
5.05
5.00
4.95
4.90
4.85
4.80
4.75
4.70
4.65
4.60
1 2 3 4 5 6 7 8 9
N (= 1/CODE RATE r)

Fig. 6 (contd)

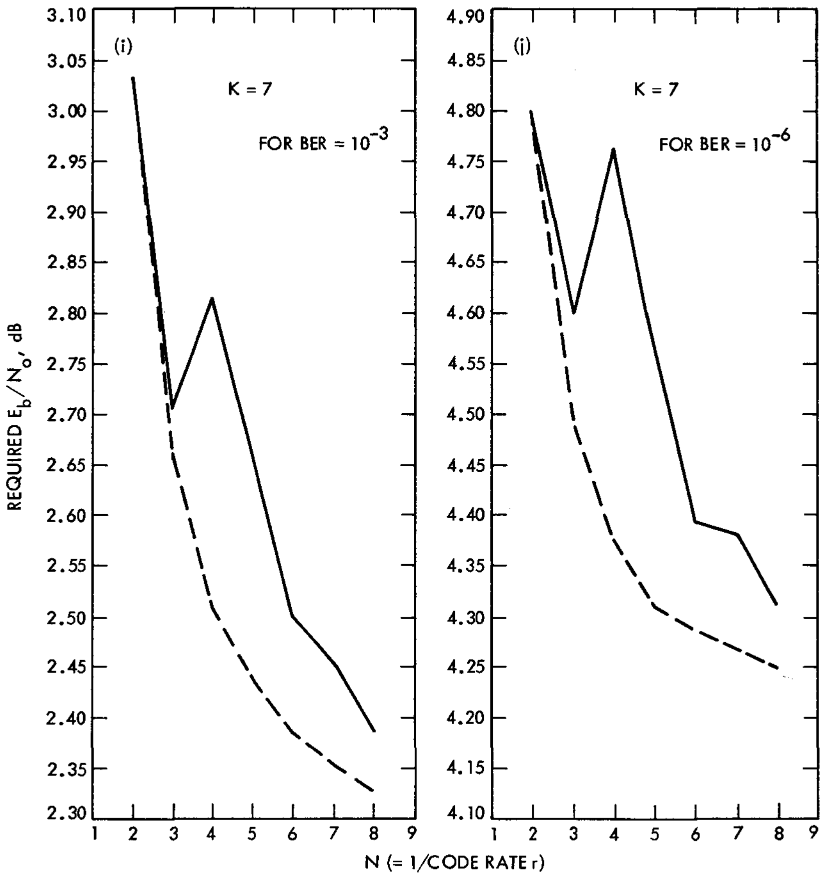


Fig. 6 (contd)

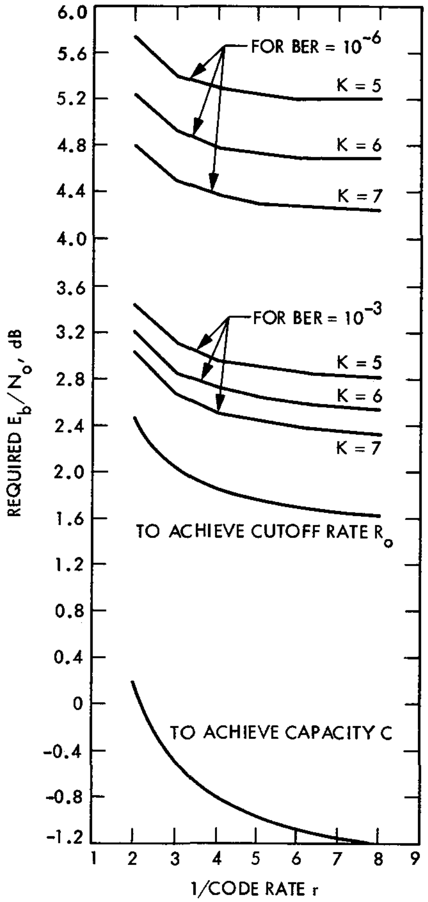


Fig. 7. Required E_b/N_o versus 1/code rate r

An Easy-to-Implement Coding Scheme for Multifrequency PPM

R. J. McEliece and L. Swanson
Communications Systems Research Section

In implementing multifrequency PPM, a naturally arising question is: Let P be a fixed number; among all integer valued processes $X_1, X_2, X_3, \ldots$ *with* $E(|X_{n+1} - X_n|^2) \leqslant P$, *which has the largest entropy? Earlier work by McEliece and Rodemich answered this question, but there is no obvious way to use this process to implement a code for multifrequency PPM. The present article describes an easy-to-implement process* $X_1, X_2, \ldots,$ *with* $E(|X_{n+1} - X_n|^2) \leqslant P$, *whose entropy is nearly as great as that of the McEliece-Rodemich process.*

I. Introduction

McEliece and Rodemich, in a study of the feasibility of using multifrequency PPM for optical communication (Ref. 1), were led to consider the following problem. Among all stationary random processes $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ taking values in the set $\{1, 2, \ldots, N\}$ and subject to the constraint

$$E(|X_n - X_{n+1}|^2) \leqslant P \qquad (1)$$

how large can the entropy be? McEliece and Rodemich showed that the maximum entropy is achieved by a Markov chain and gave an exact formula for the transition probabilities of the maximizing chain (Ref. 1). In order to apply these results to the original problem in optical communication, however, it would be necessary to encode a given binary data stream into a sequence of symbols from $\{1,2, \ldots, N\}$ which closely resembles a "typical sequence" from the optimal Markov chain. Given the complex form of the solution given in Ref. 1, such an encoding would very likely be difficult to implement. In this article we shall examine some approximations to the solution given in Ref. 1 which are not far from optimal (about 0.25 bit) and which suggest practical ways to implement energy-efficient multicolor PPM.

II. Approximations

In this section we attempt to motivate the construction to be given in Section III, as follows. We consider a simplified version of the problem mentioned in the Introduction, viz, we enlarge the state space from $\{1,2, \ldots, N\}$ to $\mathbf{Z}$, the set of all integers. For a given value of P this will increase the maximum possible entropy, but if $P \ll N^2$ this increase will be small. In any case our object in this section is to compare the entropy vs. P relationship for three possible Markov chains on $\mathbf{Z}$ subject to Eq. (1): the optimal Markov chain and two others. In Section III we will modify one of the suboptimal chains discussed here to devise a practical modulation scheme for multicolor optical communication, whose performance is (in a sense to be given below) within 0.25 bits of the theoretical optimum.

With the state set $\mathbf{Z}$, the maximum entropy Markov chain subject to Eq. (1) is a Markov chain for which the increments

$$\Delta_n = X_{n+1} - X_n \qquad (2)$$

are independent and identically distributed. The entropy of the resulting chain is just the entropy of the random variable

Δ_n, and so to solve the original problem on **Z** we just need to maximize $H(\Delta)$ subject to the condition $E(\Delta^2) \leqslant P$. This can be done using straightforward variational techniques (Ref. 2, Problem 1.8), and the maximum entropy is given parametrically as follows:

For each $\lambda > 0$, define

$$m(\lambda) = \sum_{n=-\infty}^{\infty} e^{-n^2\lambda} \tag{3}$$

so that

$$-m'(\lambda) = \sum_{n=-\infty}^{\infty} n^2 e^{-n^2\lambda} \tag{4}$$

Then among all random variables satisfying $E(\Delta^2) \leqslant P$, the maximum entropy, $H_{\max}$, is given by

$$H_{\max} = \log m(\lambda) - \lambda m'(\lambda)/m(\lambda) \tag{5}$$

where

$$P = -\frac{m'(\lambda)}{m(\lambda)} \tag{6}$$

For small values of λ the sums in Eqs. (3) and (4) are well approximated by the following integrals:

$$m(\lambda) \approx \int_{-\infty}^{\infty} e^{-n^2\lambda}\, dx = \sqrt{\pi/\lambda} \tag{7}$$

$$-m'(\lambda) \approx \int_{-\infty}^{\infty} x^2 e^{-n^2\lambda}\, dx = \frac{1}{2\lambda}\sqrt{\pi/\lambda} \tag{8}$$

Using these approximations in Eqs. (5) and (6), we obtain the approximation

$$H_{\max} \approx \frac{1}{2}\log P + \frac{1}{2}\log 2\pi e, \quad P \text{ Large} \tag{9}$$

which is extremely good even for small values of P, as exhibited in Table 1. In Table 1 we have tabulated, for a range of λ's, the corresponding value of P calculated from Eq. (6), the exact value of $H_{\max}$ calculated from Eq. (5), and the approximate value of $H_{\max}$ from Eq. (9). We conclude that for $P \geqslant 1$, there is no significant difference between the exact value of $H_{\max}$ given by Eq. (5) and the approximation given by Eq. (9).

The optimal distribution Δ_n is unfortunately not well suited for adaptation to a practical modulation scheme. The exact distribution is in fact

$$Pr\,\{\Delta = k\} = e^{-n^2\lambda}/m(\lambda) \tag{10}$$

a nonuniform distribution on a countable set $\{0, \pm 1, \pm 2, \ldots\}$ of values. However, we can get a surprisingly large entropy by considering instead of Eq. (10) a much simpler random variable $\Delta^{(L)}$, which is uniformly distributed on $\{-L, -L+1, \ldots, L-1\}$:

$$\begin{aligned} Pr\,\{\Delta^{(L)} = K\} &= \frac{1}{2L} \quad \text{if } -L \leqslant K \leqslant L-1 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{11}$$

For the sequence $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ whose increments $X_{n+1} - X_n$ are i.i.d. with common distribution $\Delta^{(L)}$, a simple calculation gives

$$\begin{aligned} P &= E\,(X_{n+1} - X_n)^2 \\ &= E\,(\Delta^{(L)\,2}) \\ &= \frac{L^2}{3} + \frac{1}{6} \end{aligned} \tag{12}$$

$$H = \log(2L) \tag{13}$$

Thus for this particular Markov chain, the relationship between the entropy H and the "power" P is

$$H = \frac{1}{2}\log P + \frac{1}{2}\log 12 + \frac{1}{2}\log\left(1 - \frac{1}{6P}\right) \tag{14}$$

Comparing Eqs. (9) and (14), we see that the difference in entropy between the optimal distribution of increments (Eq. 10) and the suboptimal distribution (Eq. 11) is approximately $1/2 \log \pi e/6 = 0.255$ bits.

As a comparison, we consider the Markov chain $\{\ldots, X_{-1}, X_0, X_1 \ldots\}$ in which the *components* X_i are i.i.d., uniformly distributed on $\{1, 2, \ldots, L\}$. In this case it is easy to calculate [cf. Eqs. (12) and (13)],

$$P = \frac{L^2}{6} - \frac{1}{6} \tag{15}$$

$$H = \log L \tag{16}$$

from which follows

$$H = \frac{1}{2}\log P + \frac{1}{2}\log 6 + \frac{1}{12P} + 0\,(P^{-2}) \qquad (17)$$

Comparing Eqs. (17) and (14), we see that for a given value of P, a Markov chain with uniform and independent increments gives one-half bit more entropy than one whose components are independent and uniform. For a fixed value of H, the distribution (Eq. 11) requires about 1.53 dB more power than the optimal distribution, whereas a uniform distribution on the X's requires 4.53 dB more power.

Motivated by these results, in the next section we introduce an encoding process which maps a sequence of 0's and 1's into a sequence of elements from $\{-L, -L+1, \ldots, L-1\}$ that closely resembles a typical sequence from Markov chain whose increments are described by Eq. (11).

III. An Encoding Algorithm

Motivated by the results of Section II, we propose a method of encoding a random stream of 0's and 1's, say U_1, U_2, $U_3, \ldots$ into a sequence $X_1, X_2, \ldots$ of elements from the set $\{-N/2, -N/2+1, \ldots, 0, 1, \ldots, N/2-1\}$, such that Eq. (1) is satisfied.[1]

First choose L to be the largest integer such that

$$\frac{L^2}{3} + \frac{1}{6} < P \qquad (18)$$

and H to be the largest integer such that

$$2^{H-1} \leqslant L \qquad (19)$$

In our encoding algorithm the symbol X_{m+1} will be determined by X_m and the $(m+1)$-st block of H data bits, viz $[U_{Hm+1}, \ldots, U_{H(m+1)}]$. These H bits are in fact used to determine an integer Δ_{m+1} in the range $[-2^{H-1}, 2^{H-1}-1]$ via two's complement arithmetic. For example,

$$\Delta_1 = \sum_{k=1}^{H} U_k\, 2^{H-k} - U_1\, 2^H \qquad (20)$$

The random variables $\Delta_1\ \Delta_2, \ldots$ are i.i.d., each being uniformly distributed between -2^{H-1} and $2^{H-1}-1$. Hence if we define the Markov chain $\{X'_m\}_{m\geqslant 0}$ by

$$\begin{aligned} X'_0 &= 0 \\ X'_{m+1} &= X'_m + \Delta_{m+1} \end{aligned} \qquad (21)$$

it follows from the results of Section II that the resulting Markov chain has [cf. Eq. (12)]

$$\begin{aligned} E(|X'_m - X'_{m+1}|^2) &= \frac{(2^{H-1})^2}{2} + \frac{1}{6} \\ &\leqslant \frac{L^2}{3} + \frac{1}{6} < P \end{aligned} \qquad (22)$$

by Eqs. (18) and (19). The entropy of $\{X'_m\}$ is H bits. Thus from Section II, if P is large, the difference between the entropy of this chain and that of the optimal Markov chain is only about 0.255 bits.

However the chain $\{X'_m\}$ is not satisfactory, since there is no guarantee that X'_m will lie between $-N/2$ and $N/2-1$. The following definition remedies the situation.

$$\begin{aligned} X_0 &= 0 \\ X_{m+1} &= \min(X_m, B) + \Delta_{m+1} && \text{if } X_m \geqslant 0 \\ &= \max(X_m, -B-1) - \Delta_{m+1} && \text{if } X_m < 0 \end{aligned} \qquad (23)$$

where

$$B = \frac{N}{6} - 2^{H-1} \qquad (24)$$

The number B defined in Eq. (24) is the largest value for X'_m that guarantees $X'_{m+1} \leqslant N/2-1$. Similarly $-B-1$ is the smallest value for X'_m that guarantees $X'_{m+1} \geqslant -N/2$. It follows that the chain $\{X_m\}$ defined by Eq. (23) will lie in the desired range $-N/2 \leqslant X_m < N/2$. The entropy of $\{X_m\}$ is still H bits, since $H(X_{m+1}|X_m) = H$ for all m. The value of $E(|X_m - X_{m+1}|^2)$ will be somewhat larger than the corresponding value for $\{X'_m\}$, since when $X_m > B$ or $X_m < -B-1$ the difference $X_{m+1} - X_m$ will no longer be uniformly distributed on $\{-2^{H-1}, \ldots, 2^{H-1}-1\}$. However, since $E(\Delta_m) = -1/2$ the rule [Eq. (23)] causes the chain to be attracted to 0, and

---

[1] For notational convenience, in this section we assume N is even, and let the state set be as described, instead of $\{1, 2, \ldots, N\}$.

unlikely to lie near the boundaries. In the Appendix we make this precise and show that in fact

$$E(|X_{m+1} - X_m|^2) \leqslant \frac{2^{2H}}{12}(1 + r^B) + \frac{1}{6} \tag{25}$$

where r is the unique solution in (0,1) to the equation

$$\sum_{k=-2^{H-1}+1}^{2^{H-1}} z^k = 2^H \tag{26}$$

Table 2 gives the value of these roots for $H = 1,2, \ldots, 7$. It follows that for a fixed P, if N is sufficiently large, Eq. (1) will be satisified by the Markov chain $\{X_m\}$.

We conclude with a simple numerical example.

Example: Let $N = 64$, $P = 23$. From Eqs. (18), (19), (24) we have $L = 8$, $H = 4$, $B = 24$. The data sequence $\mathbf{u} = (1000\ 0110\ 0111\ 0110\ 1100 \ldots)$ yields the increment sequence $\Delta_1 = -8$, $\Delta_2 = 6$, $\Delta_3 = 7$, $\Delta_4 = 6$, $\Delta_5 = -4$, and so by Eq. (23) we have $X_0 = 0$ and $(X_1, X_2, \ldots) = (-8, -14, -21, -27, -21, \ldots)$. The entropy of this chain is $H = 4$ bits, and from Eq. (25) and Table 2

$$E(|X_{m+1} - X_m|^2) \leqslant \frac{64}{3}(1 + (0.953817)^{24}) + \frac{1}{6} = 28.36$$

In fact, an exact calculation of the steady state probabilities for this chain shows that $P = 22.6$. As comparison, we note that Eq. (9) shows that the largest possible entropy for a Markov chain with $P = 22.6$ is 4.3 bits. The performance of our algorithm in this example is very close to the 0.255 bit loss predicted in Section II.

References

1. McEliece, Robert J., and Rodemich, Eugene, "A Maximum Entropy Markov Chain," *Proc. 17th Annual Conference on Information Sciences and Systems*, Johns Hopkins University, Baltimore, MD (1983), pp. 245-248.
2. McEliece, Robert J., *The Theory of Information and Coding, Encyclopedia of Mathematics and Its Applications*, Addison-Wesley Publishing Company, 1977.
3. Kleinrock, Leonard, *Queuing Systems*, Volume 2: *Computer Applications*, John Wiley and Sons, Inc., 1976.

Table 1. Comparison of H_{max} to its approximation

λ	P	H_{max} (exact)	H_{max} [from Eq. (9)]
1.00	0.499	1.0715	1.0713
0.95	0.526	1.0974	1.0974
0.90	0.5551	1.1247	1.1247
0.85	0.5880	1.1534	1.1534
0.80	0.6249	1.1838	1.1838

Table 2. Roots of Eq. (26)

H	r
2	0.414214
3	0.823408
4	0.953817
5	0.988325
6	0.997073
7	0.999268

Appendix

This appendix uses Kingman's bound, below, to bound the value of $E(X_{m+1} - X_m)^2$, where $X_1, X_2, \ldots$ is the process of Section III. This is possible because Kingman's bound gives a maximum possible value to $P(X_m \geqslant \ell)$ for $|\ell| > B$, while we have

$$E\left((X_{m+1} - X_m)^2 \mid |\ell| \leqslant B\right) = \frac{2^{2H-2}}{3} + 1/6$$

Throughout this appendix, K will mean 2^{H-1}, and N is an even integer larger than $2K$.

Lemma 1: Let $Y_1, Y_2, \ldots$ be i.i.d. random variables, $P(Y_n = \ell) = 1/2K$, $-K \leqslant \ell < K$. Let $W_0 = 0$, and let $W_{n+1} = \max(0, W_n + Y_{n+1})$. Let r, $0 < r < 1$, be a root of

$$\sum_{j=-K+1}^{K} z^j - 2K$$

Then $P(X_m \geqslant \ell) \leqslant r^\ell$. Lemma 1 is a special case of Kingman's bound (Ref. 3, p. 44).

Lemma 2: Let $Y_1, Y_2, \ldots$ be as in lemma 1. Let $T_0 = 0$ and $B = N/2 - K$, and define $T_1, T_2, \ldots$ by

$$\begin{aligned} T_{n+1} &= \min(T_n, B) + Y_{n+1} && \text{if } T_n \geqslant 0 \\ &= \max(T_n, -B - 1) - Y_{n+1} && \text{if } T_n < 0 \end{aligned}$$

Let r, again, satisfy $0 < r < 1$ and

$$\frac{1}{2K} \sum_{j=-K+1}^{K} r^j = 1$$

Then for all n, and all ℓ, we have $P(T_n \geqslant \ell) \leqslant r^\ell$.

Proof: The statement is trivial for $\ell \leqslant 0$. For any sequence $Y_1, Y_2, \ldots, T_n \leqslant W_n$, where W_n is the process of lemma 1, and so the statement is true for $\ell \geqslant 0$.

Theorem: For the process $X_1, X_2, \ldots$ described in Section III, the steady state probability P satisfies, for $\ell \geqslant B$,

$$P(X_m \geqslant \ell) \leqslant \frac{(N/2 - \ell)}{K} \cdot r^B$$

where $0 < r < 1$ and

$$\frac{1}{2K} \sum_{j=-K+1}^{K} r^j = 1$$

Proof: Given that $X_0 = 0$, $X_1, X_2, \ldots$ has exactly the same distribution as $T_1, T_2, \ldots$ in lemma 2. Thus $P(X_m \geqslant B) \leqslant r^B$.

Separately, for $B \leqslant \ell < N/2$,

$$P(X_{m+1} = \ell) = \frac{1}{2K} \sum_{j=\ell-K+1}^{N/2-1} P(X_m = j)$$

So

$$P(X_m = B) \geqslant P(X_m = B + 1) \geqslant \ldots \geqslant P(X_m = N/2 - 1)$$

Therefore

$$P(X_m \geqslant \ell) \leqslant \frac{(N/2 - \ell)}{K} \cdot r^B$$

Corollary: $E(X_{m+1} - X_m)^2 < K^2/3 + 1/6 + (4K^2/3)\, r^B$.

Proof: As shown in Section II,

$$E\left((X_{m+1} - X_m)^2 \mid -B - 1 \leqslant X_m \leqslant B\right) = \frac{K^2}{3} + 1/6$$

Therefore

$$\begin{aligned} E(X_{m+1} - X_m)^2 &< \frac{K^2}{3} + \frac{1}{6} \\ &\quad + 2 \sum_{\ell=B}^{N/2-1} P(X_m = \ell) \\ &\quad \times E\left((X_{m+1} - X_m)^2 \mid X_m = \ell\right) \end{aligned}$$

But

$$2 \sum_{\ell=B}^{N/2-1} P(X_m=\ell)\, E\left((X_{m+1} - X_m)^2 \,|X_m = \ell\right)$$

$$\leqslant 2 \sum_{\ell=B}^{N/2-1} \frac{r^B}{K} \cdot \left(\frac{1}{2K} \sum_{i=B-K}^{N/2-1} (i - \ell)^2 \right)$$

$$= \frac{r^B}{K^2} \sum_{\ell=B}^{N/2-1} \sum_{i=B-K}^{N/2-1} (i - \ell)^2$$

$$= \frac{(2K + 1)(2K - 1)}{3}\, r^B$$

$$< \frac{4K^2}{3}\, r^B$$

DSS 13 Microprocessor Antenna Controller

R. M. Gosline
Radio Frequency and Microwave Subsystems Section

A microprocessor based antenna controller system developed as part of the unattended station project for DSS 13 is described. Both the hardware and software top level designs are presented with a disucssion of the major problems encountered. Developments useful to related projects include a JPL standard 15 line interface using a single board computer, a general purpose parser, a fast floating point to ASCII conversion technique, and experience gained in using off board floating point processors with the 8080 CPU.

I. Introduction

The design of a microprocessor antenna controller was performed as part of the DSN unattended station development at DSS-13. The objectives were to replace the existing antenna controller (a Modcomp II) with a low cost standard 8 bit microprocessor controller similar to the microprocessor controllers used in the other automated subsystems, and to replace the current assembly language software with a high level programming language utilizing a top-down structured design. Another objective was to replace the antenna data collection system with a similar design.

II. System Description

A. Hardware

The system consists of two controllers as shown in Fig. 1. The controllers utilize only commercially available modules to simplify requirements for documentation and spares. Collecting analog and digital data from the various remote transducers and sending this data to the main controller located in the control room is the function of a data collection controller located in the antenna pedestal. The main controller collects data from local monitor points in the control room, provides operator interface, communication, display, and necessary logic and calculations for controlling the antenna. Communication to the station controller is through a Star Switch utilizing a JPL standard 15 line interface.

The remote data collection controller configuration is shown in Fig. 2. It consists of a single board computer (8080 CPU) with 8 K-bytes of onboard PROM and 4 K-bytes of RAM. The CPU drives a short haul modem through a serial port and is connected to 24 digital channels via onboard parallel ports. An external digital interface uses opto-isolator modules to convert 24 volt antenna levels to T.T.L. levels. A 32 channel differential input multiplexer and analog to digital converter provide interface to the analog channels. All input lines are protected from lightning discharges by filters as well as varistor type surge protectors. The following development modules, RAM module, floppy disk controller, communication expansion module, and dual floppy disk drives allowed software development and testing in the remote data collection controller, thus eliminating the need to use costly cross compilers and/or development systems. After checkout the program was burned into PROM and these development modules were removed, leaving only the CPU and analog to digital converter cards.

The main controller configuration is shown in Fig. 3. Two CPUs are utilized on the same Multibus. CPU 1 provides interface to the angle encoders through its parallel ports and performs most of the tracking related calculations. CPU 2 provides interface to the Star Switch through its parallel ports and the serial data from the data collection controller and modem through its serial port. CPU 2 performs the conversion of data to floating point and ASCII. It also handles the display and operator interface. In addition, CPU 2 performs the major part of the tasking logic and initiates appropriate action in the event of a fault detection.

Analog and digital interfaces are provided in a similar manner as the remote data collection controller, except that an input-output expansion module is necessary since the parallel ports of both CPUs are utilized. Numerical computations are facilitated by three floating point processors dedicated to tracking computations, digital servo-loop computations, and data conversion tasks such as binary to floating point or ASCII. A video-keyboard module provides interface to the local maintenance operator keyboard and high resolution monitor. Near real time status and performance data are displayed in the upper portion of the maintenance screen while response messages and operator inputs are displayed in the lower portion of the screen. Upper screen data that is not within expected limits, or states, is displayed for the maintenance operator in reverse video and flagged within the program for possible automatic monitor action. A communications expansion module provides interface to a printer for maintenance data logging, or for making listings of the program during development. It also provides an interface for an additional terminal which allows separate control of each CPU for debugging purposes. A digital to analog converter module provides the output drive voltages to operate the antenna servo-hydraulic system. As with the remote data collection controller, software was developed in the controller using a floppy disk controller and two disk drives.

B. Software

The program is written in the PL/M (an Intel trademark) language. Currently the main controller program contains about 10,000 lines of source code with 200 lines of assembly language for high speed peripheral drivers. The compiled program requires 54 K-bytes including data storage. For comparison, the minicomputer program contained 2,300 lines of FORTRAN and 4,600 lines of assembly language requiring 30 K-words of memory (a Modcomp word is 2 bytes).

The top level flow chart for the remote data collection controller is shown in Fig. 4. All of the flowcharts shown have been simplified for illustration purposes, and some modules may actually represent a combination of several. After an initialization module, the program enters an endless loop containing the routines REMDAT, which inputs data from the transducers, and REMSEND, which formats, adds a checksum, and sends the data over the serial link to the main controller. A timer on board the CPU provides a 50 ms interrupt which is divided modulo 4 so that the main loop is executed every 200 ms. The serial line operates at 9,600 baud. The data is formatted in two 73 character blocks of ASCII characters which can be displayed directly on a local maintenance terminal if desired for interface verification.

The main controller top level software is shown in Fig. 5. Both CPUs begin execution of the same code from exactly the same memory location. Both have the ability to interrogate themselves to find out which CPU they are and jump to their separate initialization routines. From there each enters a separate endless background loop except for the servicing of their interrupts.

The level 2 flowchart for CPU 1 is shown in Fig. 6. The background modules perform the following functions:

CHKSUM	Verifies that the code has not changed
SCHEDULER	Looks at current time and initiates any tasks that may be scheduled
CHKSTAK	Generates a warning message if the stack is getting too large

There are three interrupts used with CPU 1. Interrupts 3 and 4 are driven from the station clock at rates of 50 Hz and 1 Hz, respectively, which control all the time dependent motion of the antenna. Interrupt 2 is used to read a character of serial time from the station clock. The station clock is read initially to establish time. Thereafter the antenna clock is incremented in the 1 Hz interrupt routine. The 50 Hz routine counts each interrupt to insure exact synchronization with the 1 Hz interrupt. If the count is not correct or if the antenna time does not match station time, an error message is generated. The count is also used modulo 5 to call the following modules every 100 ms:

READOUTS	Inputs current position from the angle encoders
UPDATE	Updates the angle commands to the current time
SERVO	Closes the digital servo loop (type 1 or type 2 depending on start up conditions)
D2AOUT	Updates the current output drive voltage

After the interrupt 4 routine (1 Hz) steps the clock, the following modules are called:

ANAACTCNT	Decrements the counters which delay action on analog channels until after specified delays
TIMERS	Decrements action timers
CMPRATES	Computes the angle rates by differencing the position readouts

There are three tracking modes and an idle mode. An azimuth elevation mode allows pointing to a given azimuth elevation angle pair. A sidereal rate mode is used for tracking objects at sidereal rate. A three day fit mode is used for tracking objects such as spacecraft that are near sidereal rate and uses a second order fit of three values of right ascension and declination. Depending on the mode, one of the routines AEPRED, SID, or FIT3 is called to compute the current pointing angle in azimuth elevation coordinates. Three more routines are called:

CHKPRELIM	Checks the physical movement limits of the antenna and stops motion if they are about to be exceeded
REFC	Computes a refraction correction using the corrected Berman model
WINDAVG	Calculates the average wind based on the last 2 minutes; if the average wind exceeds its limit the antenna will automatically stow

The CPU 2 level 2 software is shown in Fig. 7. The background routines perform the following functions:

SPEEDSELECT	Selects either high or low speed mode depending on the distance to the desired point. An azimuth distance greater than 10 degrees or elevation greater than 3 degrees will select high speed.
CONVERT	Inputs the control room monitor channels and converts all channels (including those from the data collection controller) to floating point and ASCII. This module also checks limits and writes data to the DATAOUT buffers. If monitor action is required, flags are set for the TASKREQ module.
MONACT	Performs the actual tasks selected by the logic in TASKREQ as the result of a monitor action.
TSKACT	Performs the actual tasks selected by the logic in TASKREQ as the result of an operator or scheduler input.
DATAOUT	Outputs data from the screen, Star Switch or printer buffers.

Four interrupt routines handle the receipt of a character from the data collection controller, a Star Switch message, an operator keyboard character, or a Star Switch message time out.

Five help menus are available at the local display, the station controller display, or the remote display in the NOCC to provide the operator with assistance in remembering the commands available and their syntax. They are shown in Fig. 8. All operator and monitor events are logged on the station printer as well as reported through the Star Switch for eventual recording on magnetic media. A typical printout of a Pioneer 8 track is shown in Fig. 9.

III. Related Applications

This project produced several results that have application to other projects. A single board computer (available from Intel or National) was configured to function as a JPL standard 15 line interface port. The module and software are applicable to any Multibus microprocessor and require no external circuitry other than a cable adapter from a ribbon to an MS type connector. This module is currently being installed in transmitter and maser controllers replacing the JPL designed quad 15 line interface requiring two Multibus slots.

A general purpose parser was written to handle multiple line inputs of fields containing alphanumeric, integer, and fixed point values. The design accepts any nonalphabetic or nonnumeric character as a delimiter and considers consecutive delimiters as a single delimiter, thus providing free form inputs. An ampersand character (&) is used to indicate the end of a physical line and the continuation of a logical line for long line inputs.

Because of the requirement to update the local display with angles and analog data in near real time, a technique for fast conversion of floating point numbers to ASCII was developed. The technique is to separate the integer and fraction parts with the floating point processor into two 10 bit integers (the fractional part is scaled by 1,000). An assembly language routine then does two table look ups, using five bits at a time, to get 4 bit binary coded decimal values. Use is made of the 8080 BCD addition capability to combine the result. Conversion to

ASCII is then simply done by addition of a hexadecimal 30 to each BCD character. A floating point angle can be converted in about 300 microseconds. A faster technique, of course, is to simply use a table look up on all 10 bits. This was the original design before memory constraints required a different technique.

Although not completed at this time, two features are being included to be used with the SETI project for performing special scanning requirements. One is a raster scan pattern specified by overall size and scan rates. The second is a general three point fit capability to allow fitting a second order curve to any three points of right ascension-declination or azimuth-elevation. The current three day fit requires the points at zero hours on successive days. The more general method will allow the points to occur at unequal time intervals down to one second.

A great deal of experience has been gained configuring a maximum loaded 8080 based system with multiple processors. Difficulty was encountered in debugging problems relating to critical timing relationships between the two CPUs. Present generation logic analyzers lack the capability of triggering on combined events in each CPU and back-tracing either CPU or the bus.

Another inconvenience is the inability of the 8080 (and most other CPUs including 16 bit devices) to perform bit operations. The least addressable element is a byte. This means that to change a single bit in memory, an input-output port, or interrupt mask requires reading the byte, changing the desired bit, and storing the byte. If more than one CPU or an interrupt routine attempts to alter a bit in the same byte one of the changes may not be made. The Z80 CPU does allow bit operations, which is desirable for control applications.

All software was developed in the controllers or similar microprocessor computers. While this proved convenient from the standpoint of always having an editor or compiler available, the time required for making changes in several modules, compiling, relinking, listing, and reloading typically was 4 hours or more. Much of the debugging and patching was performed at the machine code level however, as one of the advantages of PL/M is that the source statements are easily traceable to the machine level code.

Several undesirable features of the floating point processors were encountered and should be given consideration when choosing components for future systems. Stack oriented processors cannot be shared by multiple users (such as interrupt routines or other CPUs) if their state cannot be restored to the state at the time of the interrupt. The capability to save the processor stack and restore it existed, but the status register could not be restored. It would also be useful in many applications to have access to the remainder following a fixed point division for modulo operations. Another annoying feature of the floating point processors was, if given operands that were out of range, an unpredictable result was generated rather than a closest approximation. In the refraction algorithm for example, it is necessary to evaluate small powers of *e*. If the argument is near zero, a very large result is obtained instead of unity. In time critical control applications, where it is not possible to check the range of all operands before computation, a closest approximation would be desirable.

The PL/M compiler available for the 8080 does not support floating point directly. Operands must be pushed onto the floating point processor stack, an operation code supplied, the result taken off the stack and stored. The execution time for getting numbers in and out of the stack is longer than many of the floating point operations. Some calculations can be chained to minimize this overhead but clever manipulation of a stack is one of the tasks that one hopes to avoid by using a high level language in the first place. The PL/M86 compiler for the 8086 CPU using an onboard floating point processor avoids this and allows floating point operations to be written algebraic form.

IV. Current Status

The system has been operational for several months for Pioneer 8 tracks and is now being used for all DSS 13 tracking operations. All features are not complete however, and work is continuing in the following areas: The general three point fit algorithm and raster scan for SETI are currently being implemented. The scheduler module to allow scheduled tasks to occur at a specified time and the capability of reading commonly used source predicts from an on line file are not complete at this time.

V. Conclusion

While the objectives of this effort have been achieved, it is clear that this represents a data point on the maximum capability of an eight bit microprocessor based system. Limits of memory size, execution speed, module space, and general complexity were encountered. All of these limits could be corrected by use of one of the more powerful 16 bit processors that have become available since the start of this project. The software is transferable to a 16 bit processor with minor modifications.

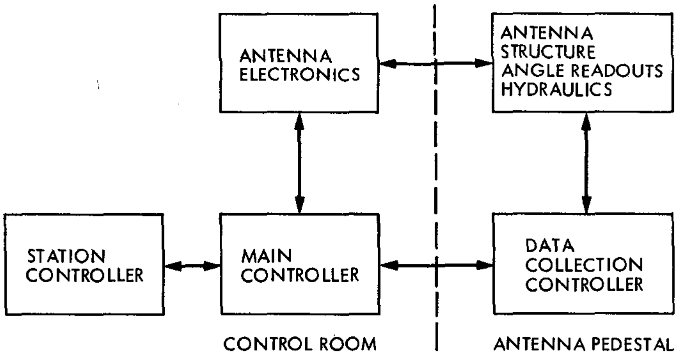


Fig. 1. DSS 13 antenna controller functional block diagram

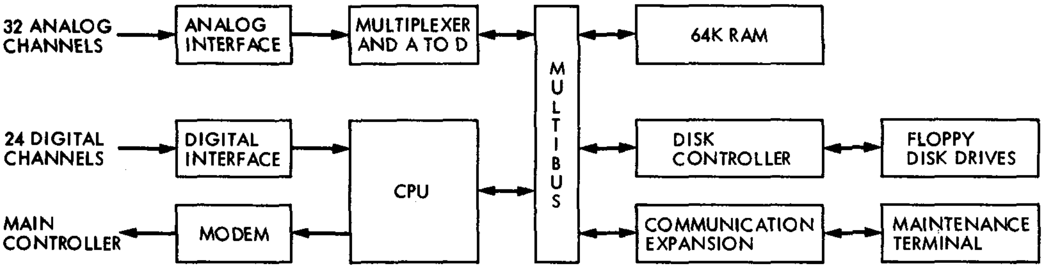


Fig. 2. Data collection controller configuration

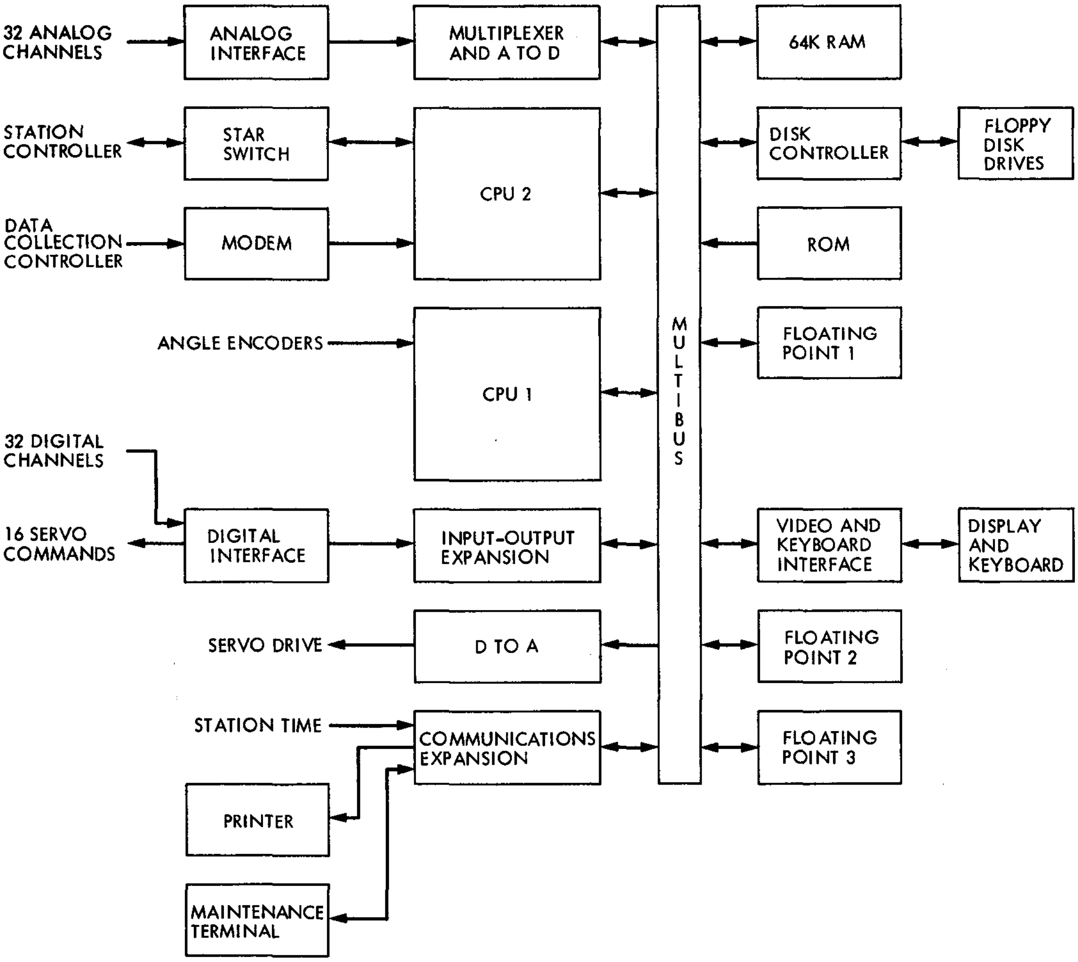


Fig. 3. Main controller configuration

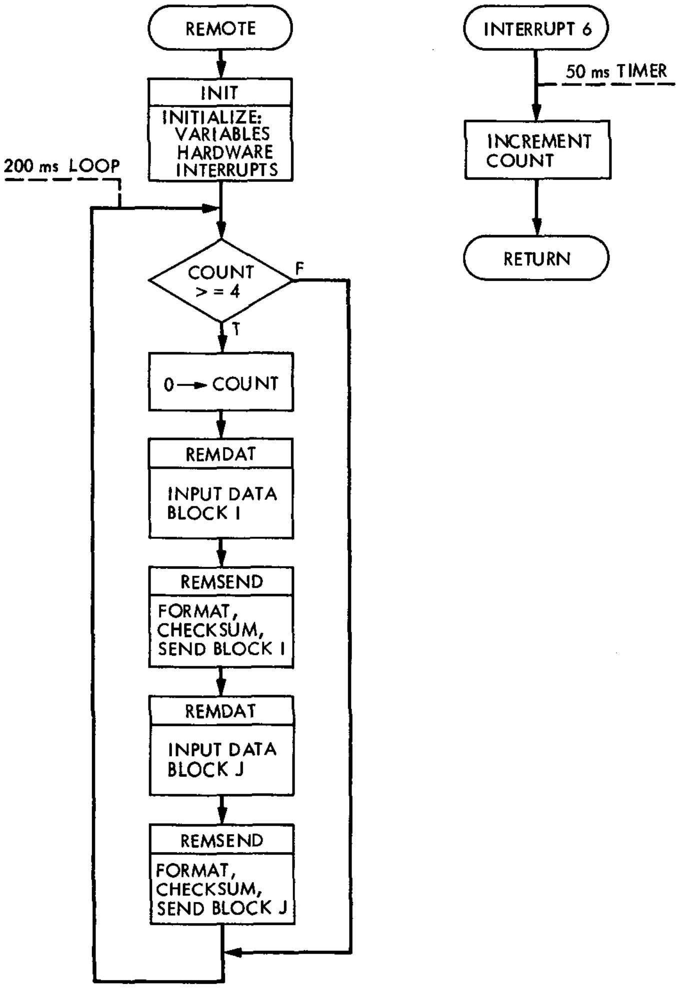


Fig. 4. Data collection controller top level software

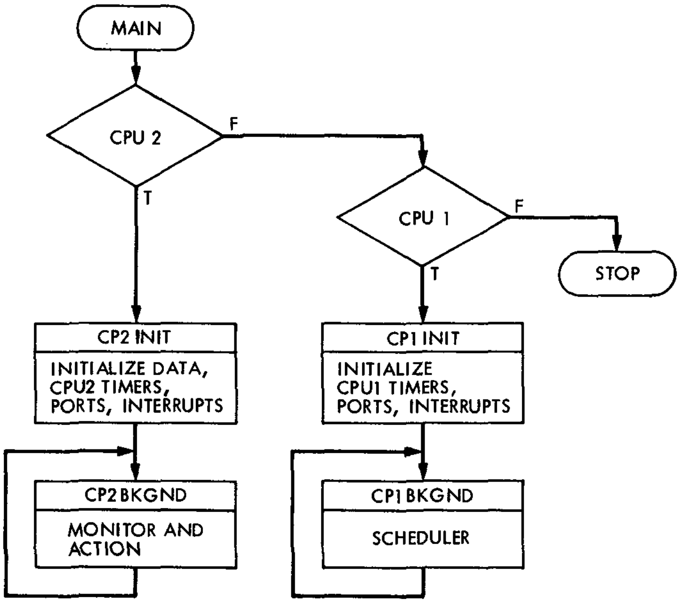


Fig. 5. Main controller top level software

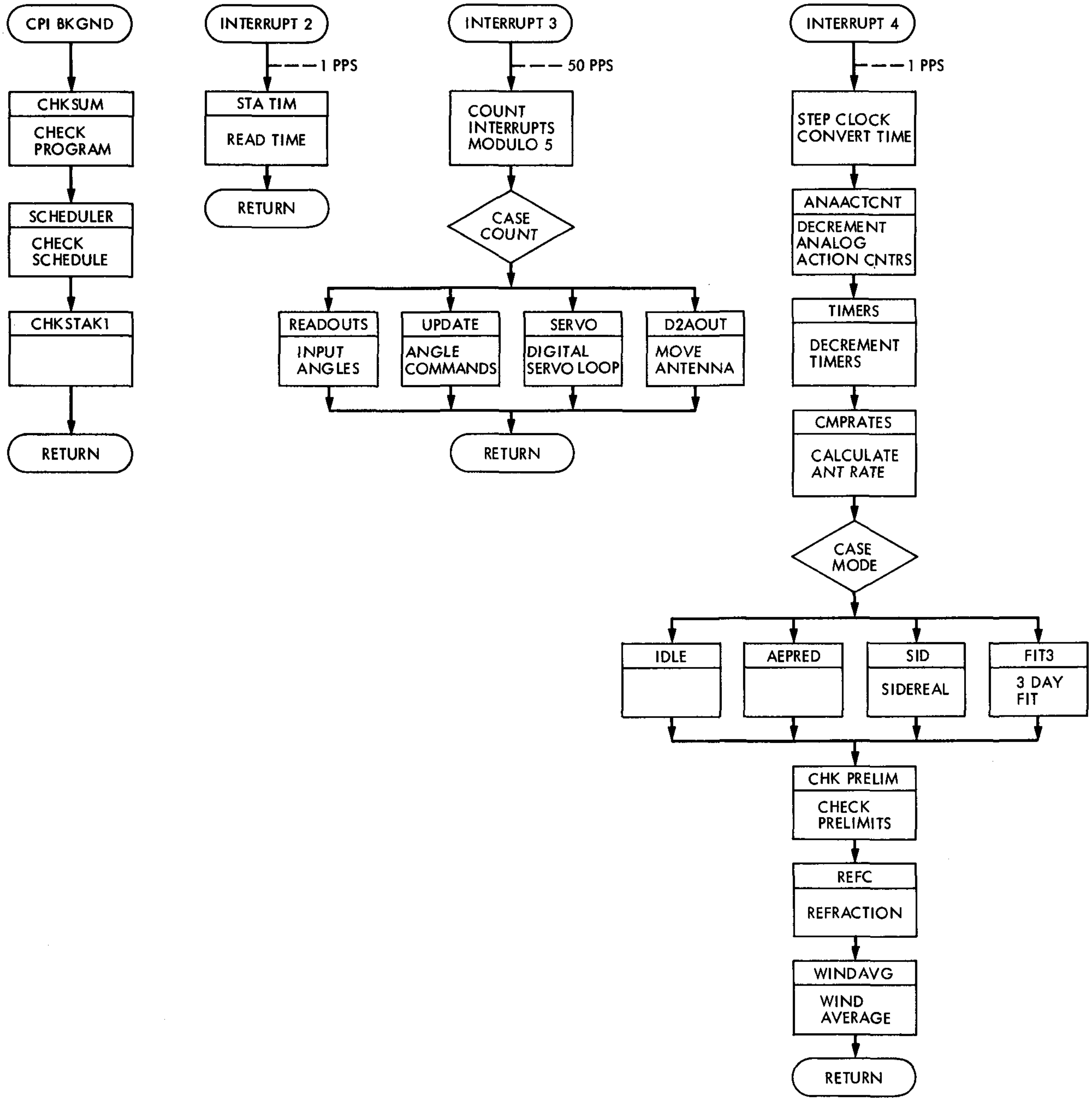


Fig. 6. CPU 1 level 2 software

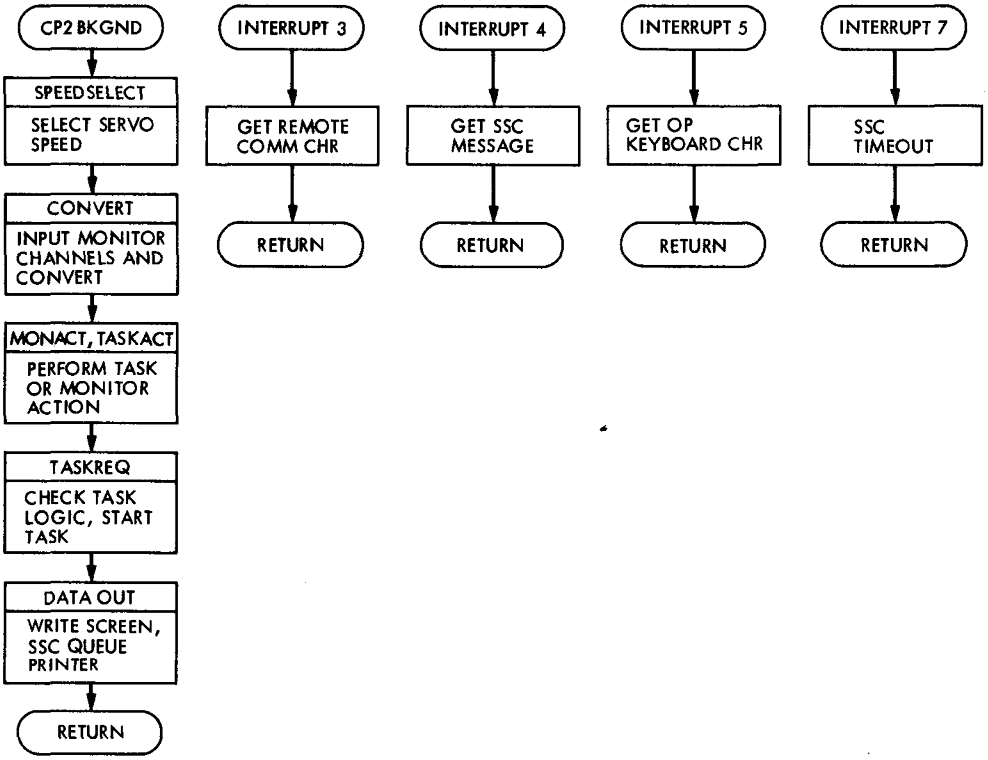
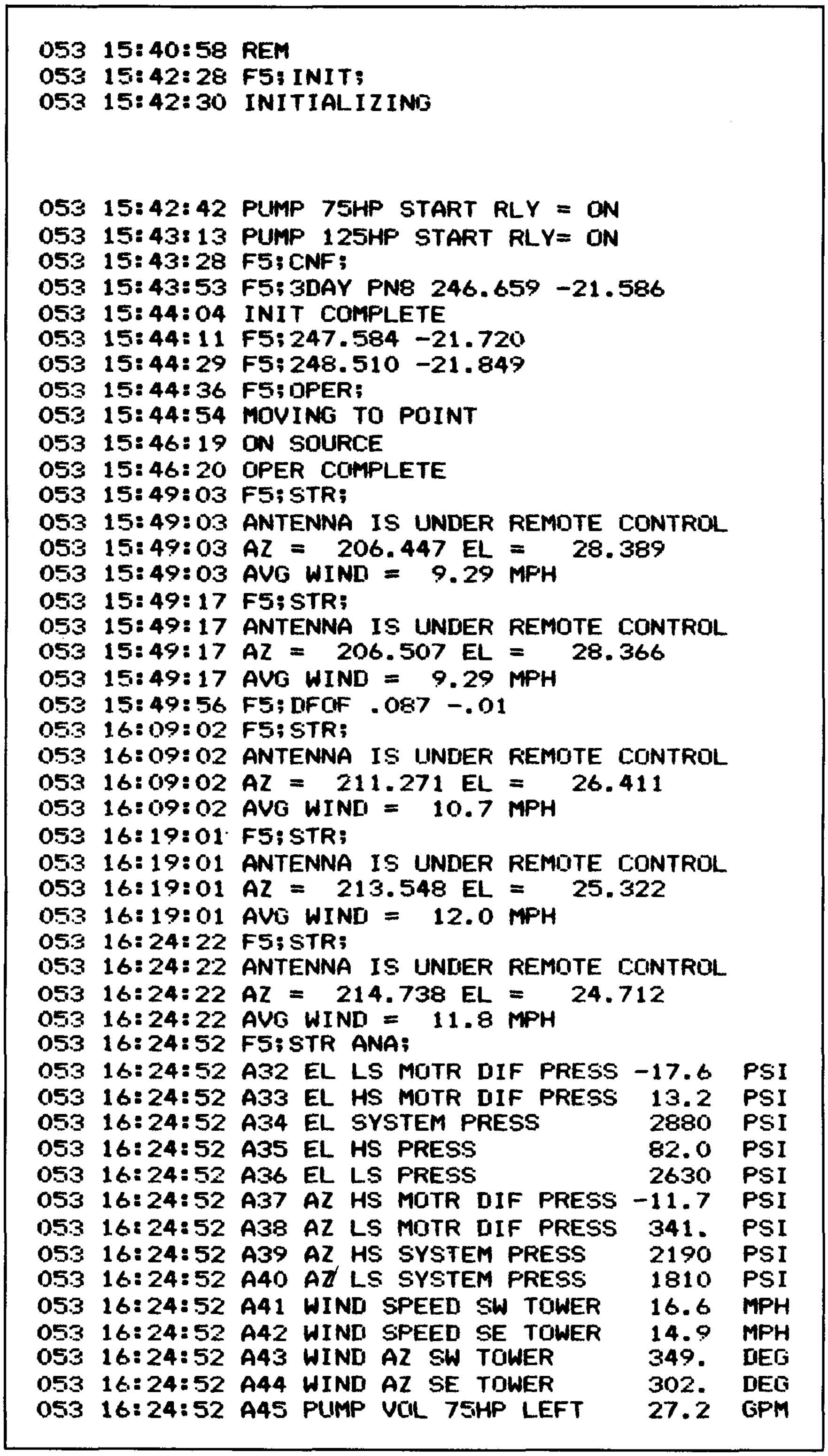


Fig. 7. CPU 2 level 2 software

```
HELP
HELP MENUS - TYPE:
 HELP 0 - station level commands
 HELP 1 - cnf formats
 HELP 2 - str formats
 HELP 3 - maint mode commands

HELP 0
ANTENNA CONTROLLER STATION LEVEL COMMANDS:
 INIT    : TURNS ON ELECTRONICS AND HYDRAULICS.
 CNF     : INDICATES START OF SCHEDULE BLOCK DATA.
 OFF     : STOPS ANTENNA, TURNS ELECTRONICS AND HYDRAULICS OFF.
 OPERATE: EXAMINE AND INITIATE ACTION SPECIFIED IN SCHEDULE BLOCK
 STOW    : ANTENNA TO STOW POSITION; ELECTRONICS, HYDRAULICS OFF.
 STOP    : SLOWS ANTENNA, APPLIES BRAKES.
 STR     : REPORT STATUS OR DATA.
 CLEAR   : RESET ANTENNA LOCKOUT AFTER A FAULT
 HELP    : DISPLAY HELP MENU

HELP 1
ANTENNA CONTROLLER CNF MENU
 SID  <name><angles>
 3DAY <name><3 sets of angles>
 AZEL <angles><wrap-up cmd>
 DFOF <angles>, [direct offsets]
 AEOF <angles>, [az/cos el]
 CLDO, [clear DFOF]
 CLPO, [clears AEOF, RDOF]
 RDOF <angles>, [ra, dec offsets]
 RAEO <angle rates>, [rate offsets]
 CLRO, [clear rate offsets]

HELP 2
ANTENNA CONTROLLER STR MENU
 STR  CNF  : CURRENT CNF BLOCK
 STR  ANA  : ANALOG DATA
 STR  DIG  : DIGITAL DATA
 STR  ANG  : ANGLE DATA
 STR  ALL  : ALL DATA
 STR  TANG : TIME, ANGLES
 STR       : SUBSYSTEM STATUS

HELP 3
MAINTENANCE COMMANDS
   10-SET BRAKE AZ       30-SET BRAKE EL
   11-RELEASE BRAKE AZ   31-RELEASE BRAKE EL
   12-HIGH SPEED AZ      32-HIGH SPEED EL
   13-LOW SPEED AZ       33-LOW SPEED EL
   14-ELECTRONICS ON     34-ELECTRONICS OFF
   15-PUMPS ON           35-PUMPS OFF
    7-INCR RATE AZ        9-INCR RATE EL
    8-DECR RATE AZ        0-DECR RATE EL
   19-IGNORE LIMITS AZ   39-IGNORE LIMITS EL
   20-SET LIMITS AZ      40-SET LIMITS EL
   21-PRINT CURRENT INCREMENTAL RATES
   23-PRINT RATE INCREMENT
```

Fig. 8. Help menus



```
053 15:40:58 REM
053 15:42:28 F5;INIT;
053 15:42:30 INITIALIZING



053 15:42:42 PUMP 75HP START RLY = ON
053 15:43:13 PUMP 125HP START RLY= ON
053 15:43:28 F5;CNF;
053 15:43:53 F5;3DAY PN8 246.659 -21.586
053 15:44:04 INIT COMPLETE
053 15:44:11 F5;247.584 -21.720
053 15:44:29 F5;248.510 -21.849
053 15:44:36 F5;OPER;
053 15:44:54 MOVING TO POINT
053 15:46:19 ON SOURCE
053 15:46:20 OPER COMPLETE
053 15:49:03 F5;STR;
053 15:49:03 ANTENNA IS UNDER REMOTE CONTROL
053 15:49:03 AZ =  206.447 EL =   28.389
053 15:49:03 AVG WIND =  9.29 MPH
053 15:49:17 F5;STR;
053 15:49:17 ANTENNA IS UNDER REMOTE CONTROL
053 15:49:17 AZ =  206.507 EL =   28.366
053 15:49:17 AVG WIND =  9.29 MPH
053 15:49:56 F5;DFOF .087 -.01
053 16:09:02 F5;STR;
053 16:09:02 ANTENNA IS UNDER REMOTE CONTROL
053 16:09:02 AZ =  211.271 EL =   26.411
053 16:09:02 AVG WIND =  10.7 MPH
053 16:19:01 F5;STR;
053 16:19:01 ANTENNA IS UNDER REMOTE CONTROL
053 16:19:01 AZ =  213.548 EL =   25.322
053 16:19:01 AVG WIND =  12.0 MPH
053 16:24:22 F5;STR;
053 16:24:22 ANTENNA IS UNDER REMOTE CONTROL
053 16:24:22 AZ =  214.738 EL =   24.712
053 16:24:22 AVG WIND =  11.8 MPH
053 16:24:52 F5;STR ANA;
053 16:24:52 A32 EL LS MOTR DIF PRESS -17.6  PSI
053 16:24:52 A33 EL HS MOTR DIF PRESS  13.2  PSI
053 16:24:52 A34 EL SYSTEM PRESS       2880  PSI
053 16:24:52 A35 EL HS PRESS           82.0  PSI
053 16:24:52 A36 EL LS PRESS           2630  PSI
053 16:24:52 A37 AZ HS MOTR DIF PRESS -11.7  PSI
053 16:24:52 A38 AZ LS MOTR DIF PRESS  341.  PSI
053 16:24:52 A39 AZ HS SYSTEM PRESS    2190  PSI
053 16:24:52 A40 AZ LS SYSTEM PRESS    1810  PSI
053 16:24:52 A41 WIND SPEED SW TOWER   16.6  MPH
053 16:24:52 A42 WIND SPEED SE TOWER   14.9  MPH
053 16:24:52 A43 WIND AZ SW TOWER      349.  DEG
053 16:24:52 A44 WIND AZ SE TOWER      302.  DEG
053 16:24:52 A45 PUMP VOL 75HP LEFT    27.2  GPM
```

Fig. 9. Typical operation record

TDA Assessment of Recommendations for Space Data System Standards

E. C. Posner and R. Stevens
Office of Telecommunications and Data Acquisition

NASA is participating in the development of international standards for space data systems. The purpose of this article is to provide a TDA assessment of the recommendations for standards thus far developed. The Recommendations for Telemetry Coding and Packet Telemetry provide worthwhile benefit to the DSN; their cost impact to the DSN will be small. They will be of advantage to the NASA space exploration program. Their adoption should be supported by TDA, JPL, and OSTDS.

I. Introduction

NASA is participating in the development of international standards for space data systems. Work on some of the system elements is mature and is being readied for concurrence by the participating space agencies.

The purpose of this article is to provide a TDA assessment of the recommendations for standards thus far developed. Recommendations for Telemetry Coding and Packet Telemetry are considered in detail; others, which are less well developed, are considered briefly.

A. Summary of Findings

The Recommendations for Telemetry Coding and Packet Telemetry provide worthwhile benefit to the DSN; their cost impact to the DSN will be small. They will be of advantage to the NASA space exploration program. Their adoption should be supported by TDA, JPL, and OSTDS.

Six other Recommendations were also reviewed. They all are less mature than the Telemetry Recommendations. They deal with standardizable elements of the system, and it is expected that they will lead to adoptable standards after the appropriate additional work.

B. Background

The Consultative Committee for Space Data Systems (CCSDS) is in the process of developing a set of Recommendations for standardization of space data system functions by space agencies. The purpose of the development is to enable economical cross-support between space agencies. From NASA's standpoint, it is expected that the Standards as issued by NASA in accordance with the Recommendations will also facilitate support of NASA's own missions by the Deep Space Network (and by TDRSS).

The CCSDS has representatives of European, Japanese, and other national space agencies, including NASA. JPL has several active members on the CCSDS Panels. The process of developing CCSDS standards Recommendations involves a sequence of increasingly mature documents: Concept Papers, White Books, Red Books, Blue Books. The details of the process are described in the CCSDS Recommendations Documents (e.g., CCSDS Recommendations for Space Data Systems Standards, Packet Telemetry, (Red Book) Issue-1, September 1983, page ii: "Document Status").

The Telemetry Coding and Packet Telemetry Recommendations are the farthest along. They are being readied for formal concurrence in March, 1984. The Coding Recommendation is based on mature technology. It is essentially the technique planned for use by Voyager 2 at Uranus.

The Packet Recommendation will produce a "leading standard." That is, its past applications to space data systems are few and have only demonstrated the concept, but its future potential is considered good. For example, a relatively simple adaptive packet telemetry system was used for the low rate channels on Seasat A with good success (Ref. 1). The system was very effective in supporting the ground data handling and distribution functions. Also, the JPL ISPM spacecraft data system design was initiated, but not completed, with a packet telemetry structure closely related to that of the present Recommendations. The JPL Mars Geoscience/Climatology Observer Project has made a preliminary assessment of using packet telemetry and has concluded that it would be cost effective (private communication).

In addition to the two Telemetry Recommendations, there are a number of others being prepared by the CCSDS Panels. They are:

(1) Time Code Formats (Draft Red, Blue Books)

(2) Packet Telecommand (White Book)

(3) Radio Frequency and Modulation (Concept Paper)

(4) Standard Format Data Units (White Book)

(5) Electronic Communications (pre-Concept Paper)

(6) Radio Metric and Orbit Determination (pre-Concept Paper)

Recommendations (2), Packet Telecommand, and (3), Electronic Communications, are expected to have the greatest impact on DSN service. However, none of these six are further discussed here.

II. Detailed Assessment of Impact of the Recommendations on the DSN

A. Summary of Approach

A block diagram of the DSN Telemetry System for the 1988-90 period was developed (see Fig. 1). It is a direct evolution of the mid-80's MK IVA; it includes the currently planned SFOC/NOC concept. In this, the current split of functions between the Mission Control and Computation Center (MC^3) and the DSN Network Operations Control Center (NOCC) is changed somewhat to reduce minor duplication, creating a Flight-Project implemented and operated Space Flight Operations Center (SFOC) and a DSN Network Operations Center (NOC). Using the diagram, the elements of the system that were seen to be affected by the Telemetry Recommendations were identified by a study team from the TDA Office.

Figure 1 depicts the DSN Telemetry System of the 1988-90 SFOC/NOC era. It encompasses that portion of the end-to-end telemetry system for which the DSN has responsibility. The diagram identifies those changes to the DSN Telemetry System that will be required to support missions using the CCSDS Telemetry Coding and Packet Telemetry Recommendations. It shows telemetry frame sync and Reed-Solomon decoding being done at the Signal Processing Centers (vs. the NOC) in a new "DSN Standard Decoder." The station location is not required by the Telemetry Recommendations, and may turn out not to be desired. In any case, the location chosen is independent of the requirements to support the standards. The Telemetry System and interface functions requiring changes related to the adoption of the Recommendations are identified in the boxes of Fig. 1.

The marginal implementation and operating costs for the DSN to adopt the Telemetry Recommendations were estimated. Also, the operating characteristics of the DSN incorporating the Telemetry Recommendations were assessed.

Finally, the suitability of the Recommendations for the JPL end-to-end space data system (Ref. 2) was considered, from a TDA perspective, by study team members who have participated in the JPL effort to develop the CCSDS Recommendations.

B. Assessment of Telemetry Coding Recommendation

1. Description. The Recommendation is CCSDS Blue Book, Draft Issue-0, February 1984. The coding is for the link between the spacecraft and the ground signal processing center. Proper coding and decoding of the link improves its performance by a factor of 5 to 7 dB at a bit error rate (BER) of 10^{-5}. The very low BER is needed to transport spacecraft science data that have been compressed and to enable effective use of packet data streams (Ref. 3).

The Recommendation proposes both an inner and an outer code. The inner code is a rate 1/2 constraint length 7 convolutional code with Viterbi decoding (Ref. 4). The outer code is a Reed-Solomon (255,223) block code with symbols from the 256-element field interleaved to depth 5 (Ref. 5). The inner code can be concatenated with the outer code, or it can be used separately as in Ref. 4. The coding is compatible with, but does not require the use of, the proposed packet telemetry concept. Also, an uncoded link is allowed.

Two changes to the original Red-Book Recommendation proposed as a result of the TDA assessment have been accepted by the CCSDS Panel, and incorporated into the draft Blue Book, upon which the discussion of this article is based. They related to standardizing on soft decision Viterbi decoding so that deep space missions can be assured of cross-support, and putting the packet sync word in the Reed-Solomon code blocks so that coding and packets are independent to aid testing and to allow independent evolution. They also foreshadowed the possible need to go to constraint length 7 rate 1/3 convolutional codes for some missions (Ref. 6).

2. **Impact on DSN.** Minor software and hardware changes are required in the Viterbi decoder, the Telemetry Processing Assembly, and their monitor and control interfaces. Also, frame sync and Reed-Solomon decoding hardware and software are required in the DSN to support the Recommendation. However, those sync and decoding functions must be provided by the DSN to support the SFOC/NOC concept for committed mission support, independent of the adoption of the standards. As we will see in a subsequent section of this article, the advantages of the Coding Recommendation outweigh the disadvantages.

Table 1 gives a summary of estimated implementation and operations costs for adopting the Telemetry Coding Recommendation. The costs involved are small ($70K). Specifically, the delta cost of providing the standard decoding capability is limited to the cost of implementing the capability to reverse the order in which the data symbols are sampled in the maximum likelihood-convolutional decoder. This single change provides compatibility with the "blue book" standard for Viterbi decoding.

Additional operational costs attributable to implementation of the Coding Recommendation are negligible, as testing and training of operators for a new decoder should be no more expensive than for maintaining the old capability. (Likewise, operational support of packet telemetry should be no more difficult to provide than with the present mission-unique interfaces.) In fact, once a standard has been adopted, testing and training should become simpler, as support will be required for only one data interface for all projects.

The Coding Recommendation is fundamentally sound. Some details will probably need modification in application. The basic technology embodied in it has had extensive development and flight/ground application. It is a well proven and effective concept for deep space missions. Concatenated convolutional/Reed-Solomon coding, in essence identical to that defined by the Recommendation, will be supported by the DSN and flight proven for deep space use by Voyager 2 at Uranus (Ref. 7). The recommended convolutional coding/decoding has been used by Voyager 1 and 2 from launch and by TDRSS; it will be used by Galileo, ISPM (ESA), Giotto (ESA), and AMPTE (Ref. 8). Galileo and Giotto will use concatenated Reed-Solomon coding very similar to that of the Recommendation (Ref. 9).

Cross support of one agency's spacecraft by another agency's tracking network (Ref. 10) will be simplified by embracing the Recommendation. That will be a benefit to the DSN in the future. Ultimately, when all spacecraft not conforming to the standards have expired, fewer code options will require sustaining and operational support by the DSN. That will reduce DSN support costs.

Although technology developments during the 90's may allow a 1 to 2 dB increase in link capability, the proposed standard is expected to have good longevity through the 90's for deep space mission use. However, it may be that the standard will be modified to permit constraint-length 7 rate 1/3 convolutional codes. This would permit a gain of 0.5 dB (Ref. 6) at little additional complexity on spacecraft and ground. Thus, the Coding Recommendation was endorsed to OSTDS as desirable for the DSN and NASA.

C. Assessment of Packet Telemetry Recommendation

1. **Description.** The Recommendation assessed was CCSDS Blue Book, Draft, Issue-0, February 1984. The packets, consisting of data and an identifying header, are made up by the spacecraft individual data sources. They are asynchronously time multiplexed into the spacecraft data stream. They pass through the system intact to the user of the data. For transfer from the spacecraft through delivery to the Project's mission control, the packets are carried in "telemetry transfer frames." The frames carry information on the mission and instructions for extracting the packets. The frames are multiplexed synchronously into the data stream. On the ground, synchronizing to and reading of the frame header allow the data stream to be directed toward the proper spacecraft Project. Then the individual packets can be demultiplexed and read by the Project (Ref. 11).

The concept enables multimission identification, extraction, and distribution of the data in the stream. It also enables but does not require an adaptive data stream that reacts to events on board the spacecraft. That capability is expected to benefit spacecraft projects and simplify ground processing. The telemetry link efficiency with a data stream compatible with the Packet Recommendation should be comparable to that of current deep space data system designs. Indeed, calculations show that the approximately 0.25 dB average link inefficiency which can be computed (details omitted) for packetized telemetry is about the same as or lower than the inefficiencies computed by the JPL Telecommunications Systems Section (private communication) for Voyager and other fixed-format time-division multiplexed telemetry.

The Recommendation as we have said defines the format for the packets and the transfer frame. The detail in the Recommendation document is sufficient to insure compatibility between users. Several options are available within the Recommendation to handle packets that are longer than a single frame permits. The Recommendation is compatible with, but does not require, per se, the use of the channel coding recommended. It does, however, require low error probability in the channel, as can be provided by proper application of the coding recommendation.

2. **Impact on DSN.** In concept, the packets, carried by their transfer frames, travel through the ground data system unobserved until they are demultiplexed from their frames at their destination. Therefore, the impact on the DSN is small.

Currently (Ref. 12), the DSN NOCC finds the telemetry frame header and does some demultiplexing to provide data required by DSN performance analysts. The simple algorithm used will not suffice to demultiplex an asynchronous packet stream. However, in the SFOC/NOC era, the DSN NOC is to receive the required information from the SFOC's demultiplexing process. This appears to be a desirable feature of the SFOC/NOC concept, particularly for support of asynchronous packetized data streams.

Full and effective use of the packet data concept requires very low error rates in the end-to-end system. The error probability requirements are two to three orders of magnitude more stringent than generally needed with previous deep space imaging data systems. It is not evident that any system changes to the DSN Complexes or the GCF will be needed solely to support the more stringent requirements.

Surely, however, more rigorous performance validation will be required. The costs have been estimated for providing automatic retransmission or forward error correction on the GCF present or future 56- to 600-kbps wideband data lines. This capability will be appropriate for real-time transmission of high-rate compressed images or packet telemetry. However, the costs ($360K) were not included as a marginal cost for adopting packet telemetry. This is because the capability would only be provided in response to a specific mission requirement and could be driven by missions not using packets, such as missions using data compression (Ref. 3). This low-error capability could also be required by a mission using packets but not in conformance with the CCSDS Recommendations.

The estimated DSN implementation and operational costs to adopt the Packet Telemetry Recommendation are also given in Table 1. The cost impact, without the GCF implementation noted above, is small ($60K). Thus, the Packet Telemetry Recommendation was endorsed to OSTDS as desirable for the DSN and NASA.

D. Considerations for Missions Not Using Telemetry Standards

There are no spacecraft currently designed or in flight that are fully compatible with the Recommendations. For their lifetimes, those spacecraft must be supported by DSN operating modes, or equipment, different from that for standardized spacecraft. Also, it should be anticipated that some future spacecraft designs, for reasons of overall economy, will use "inherited" flight equipment that will result in data systems incompatible with the standards.

For those reasons, for a long time into the future, capability for support of nonstandard spacecraft data systems will have to be retained by the DSN. Even so, adopting the standards is expected to lead to fewer, not more, telemetry designs to support during the period. The basic approach to planning for removal of no longer needed capability in the DSN will be to obtain concurrence of affected projects, both present and potential future, and of affected NASA Centers, before removal of existing capability.

E. Advantages and Disadvantages

Table 2 lists various advantages and disadvantages of the Telemetry Standards that were uncovered in the course of the TDA study. They are listed in the table without quantification. Nevertheless, it is seen that advantages far outweigh disadvantages.

III. TDA Perception of Effect of Standards for Space Data Systems

A. Introduction

This section presents the TDA assessment of how Data System Standards based on the CCSDS Recommendations will affect the way JPL and NASA conduct missions in the future. Issues of cross-support, mission design, mission operations, and space program growth are assessed in this light. The overall conclusion is that the Standards will promote JPL and NASA programs by reducing costs and increasing cooperative ventures.

B. Cross-Support

The greater interoperability that the Standards enable will make it easier for NASA to give and get cross-support. A greater number of cooperative missions between space agencies, such as the International Solar Terrestrial Physics Program (Ref. 13), can be expected. The common performance standard must be maintained high enough so that NASA can receive cross-support as well as give it. This is because NASA's deep space missions usually require performance near attainable limits.

C. Mission Design

The Standards can benefit both the U.S. and International space programs in the mission design phase. It is expected that the Packet Telemetry Standard will facilitate mission-independent instrument integration including re-use and reflights. Adaptive telemetry and data compression become easier.

D. Mission Operations

The benefits can be substantial, as the SEASAT A operations experience has shown. It will become much easier to handle data from acquisition to distribution and archiving. Common program data, e.g., planetary data system, will be easier to provide. Likewise, coordination of data from multiple spacecraft, such as the ISTP Program and Space Station, will be easier. It will become much easier to have distributed operations and science teams. Common hardware and software are expected to reduce costs in the mission operations system. Overall, the Standards will benefit mission operations.

E. Space Program Growth Potential

The international Space Data System Standards can be a strong factor in the growth of the world's space program. The Standards should cut lead time and cost of experiment integration. Multinational experiments on a space agency's mission will be easier, as well as multispacecraft multiagency missions. Cross-support will be cheaper to obtain or provide, and this in turn can encourage participation by other nations in NASA programs. The availability and utility of space-derived data will increase.

IV. Future Work to Support Standards

For the U.S. to be a full participant in and beneficiary of the international Space Data System Standards program requires a continuing effort. This effort is to evolve the Standards to take advantage of new spacecraft and ground technology and mission concepts. The possibilities for enhancing space programs require a broad planning perspective for NASA, JPL, and TDA within JPL. The TDA responsibility is to determine performance and interface specifications and verify system performance against theory and specification. One change to be encouraged early is to promote the international standardization of the common JPL/GSFC constraint-length 7 rate 1/3 convolutional code as an alternative to the constraint-length 7 rate 1/2 code of the current Guidelines. This can benefit missions which are not bandwidth constrained and can reduce costs or increase science value from the 0.5 dB increase in performance (Ref. 6).

V. Conclusions

The adoption of appropriate standards for space data systems will be of long term benefit to NASA. It will reduce the cost of mission design, implementation, operations, and data analysis. These benefits will be obtained by more thorough development of spacecraft and ground mission and facility designs and operating procedures. Joint ventures between NASA and the space agencies of other nations will be made easier, better, and less costly. Space data will be more available.

Two of the CCSDS Recommendations, Telemetry Coding and Packet Telemetry, have been concurred in by TDA at JPL. They have been well developed and have sound bases. They do not cause noticeable expense to NASA in the near term and can only save in the long term. Concurrence in them was therefore recommended to OSTDS by the TDA Office at JPL.

References

1. Brown, J. W., et al., "Seasat Low-Rate Data System," *Science*, Vol. 204 (1979), pp. 1407-1408.
2. Diaz, A. V., Marcia Neugebauer, J. Stuart, and Richard B. Miller, "New Approaches to Planetary Exploration – Spacecraft and Information Systems Design," *Proc. 1983 International Astronautics Federation General Meeting*, Budapest, Hungary, October 1983, Published by AIAA, 1983, no pp. given.
3. Pierce, John R., and Edward C. Posner, *Introduction to Communication Science and Systems*, Plenum, New York, 1980, Sec. 13.6.
4. Deutsch, L. J., and R. L. Miller, "The Effect of Viterbi Decoder Node Synchronization Losses on the Telemetry Receiving System," *TDA PR 42-68* (April 1982), pp. 17-24.
5. Divsalar, D., and J. H. Yuen, "Performance of Concatenated Reed-Solomon/Viterbi Channel Coding," *TDA PR 42-71* (November 1982), pp. 81-94.
6. Butman, S. A., L. J. Deutsch, and R. L. Miller, "Performance of Concatenated Codes for Deep Space Missions," *TDA PR 42-63* (June 1981), pp. 33-39.
7. Posner, Edward C., and Robertson Stevens, "Deep Space Communication – Past, Present, and Future," invited paper, *IEEE Communications Magazine*, Vol. 22, No. 4 (May 1984).
8. Burt, R., "Mark IV-A DSCC Telemetry System Description," *TDA PR 42-68* (April 1982), pp. 130-138.
9. MacMedan, M. L., "Standard Methods for Telemetry and Telecommand Coding," *Proc. 1983 International Telemetry Conf. (ITC '83)*, San Diego, CA, Oct. 24-27, 1983, published by International Foundation for Telemetring, Woodland Hills, CA, 1983, pp. 839-846.
10. Layland, J. W., et al., "Interagency Array Study Report," *TDA PR 42-74* (Aug. 1983), pp. 117-148.
11. Hooke, Adrian J., "Packet Telemetry and Packet Telecommand: The New Generation of Spacecraft Data Handling Techniques," *Proc. ITC '83* (*op.cit.*), pp. 829-838.
12. Hall, J. R., "Network Control System," *DSN PR Volume XII*, Oct. 1972, pp. 5-11.
13. Wardrop, M. Mitchell, "A New Plan for Space Physics," *Science*, Vol. 222 (22 Oct 1983), pp. 400-401.

Table 1. Estimated DSN marginal costs[a] for implementing Telemetry Channel Coding and Packet Telemetry Recommendations

Standards / Cost Elements	Channel Coding: Viterbi Decoding	Channel Coding: R-S Decoding[b]	Packet Telemetry[c]
Hardware implementation	47.5	0	0[d]
Software implementation	23.9	0	57.5[d]
Operations	0	0	0

[a] All costs in thousands of FY 84 dollars, ±30% estimate.

[b] Assumes use of current R-S Decoder hardware/software now available in SFOC. If decoding function were shifted to DSN SPC's in future, there would be additional hardware and training costs. However, those costs are independent of adoption of standards.

[c] Assumes fixed length transfer frames as per Recommendation.

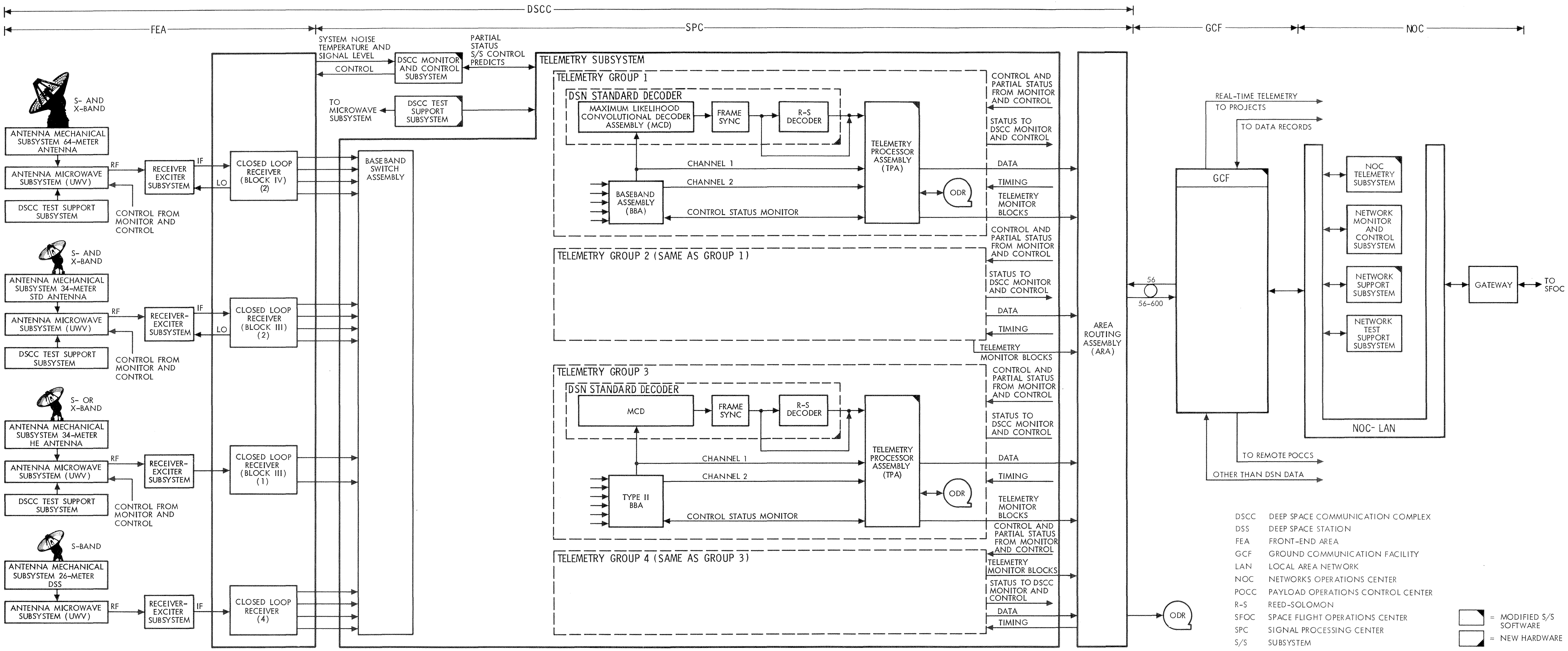
[d] Auto recall of GCF WBDL (wideband data lines): cost estimated to $360K – $193K for hardware, $167K for software. Needed only for missions requiring real-time high data rate with low deletions. This would be a mission requirement, if it arises, independent of whether the CCSDS Packet Telemetry Recommendations were adopted.

Table 2. Characteristics and effects of telemetry guidelines

Guideline	Characteristic	Effect
	Positive	
Coding	Common channel coding for all DSN-supported missions	Simplifies all phases (planning through design through operation) of DSN support of all missions (Deep Space, HEO, NASA, Non-NASA)
		Wider base, greater depth of understanding by users of coded channel capability and characteristics
	Cross-support of U.S. missions by other Agencies and conversely	More cross-support potential
	Enables full use of Packet Telemetry	More cross-support potential
	Coding proposed is essentially that evolving as a *de facto* DSN standard for deep space mission support	Builds on considerable DSN investment
		Technology sufficiently mature to support providing a stable standard
		Appropriate time frame for initial establishment of standard
	Allows mission-independent decoder software/hardware	Reduced hardware/software at Complexes
		Reduced effort to test and verify performance for mission set
		Reduced effort for sustaining in Net
	Reed-Solomon/Viterbi provides low error probability for all missions using it without bit power penalty	Low-error-probability engineering telemetry may make isolation of spacecraft vs ground anomalies easier
	Enables mission-independent frame sync	Reduced implementation and test costs and increased confidence
	Negative	
	Retention of old nonconforming designs in overlap period	Increased sustaining engineering costs
	Positive	
Packet	Enhanced ability to obtain cross-support	Reduced NASA costs if facilities of other agencies are needed
	Enhanced ability to provide cross-support	Broader base of DSN customers; more joint missions
	Makes delivery of data to Projects easier	Reduced end-to-end system test and operations costs, increased user satisfaction, and greater cross-support potential
	Eliminates potential need for multiple subcarriers	Fewer Subcarrier Demodular Assemblies and Symbol Synchronizer Assemblies needed at Signal Processing Centers

Table 2. (cont'd)

Guideline	Characteristic	Effect
	Negative	
Packet	Required packet system performance not specified	Potential user surprise and dissatisfaction
	Performance of full adaptive options not specified or understood	Potential need to implement still better coding or provide more signal-to-noise some other way
	Possible complication of project end-to-end testing involving DSN	May increase Signal Processing Center time needed for testing
	Need to retain non-packetized telemetry for old designs	Possible enlarged complexity in Telemetry Processing Assembly and increased sustaining



DSCC	DEEP SPACE COMMUNICATION COMPLEX
DSS	DEEP SPACE STATION
FEA	FRONT-END AREA
GCF	GROUND COMMUNICATION FACILITY
LAN	LOCAL AREA NETWORK
NOC	NETWORKS OPERATIONS CENTER
POCC	PAYLOAD OPERATIONS CONTROL CENTER
R-S	REED-SOLOMON
SFOC	SPACE FLIGHT OPERATIONS CENTER
SPC	SIGNAL PROCESSING CENTER
S/S	SUBSYSTEM

Fig. 1. Baseline 1988–90 era DSN Telemetry System

The Goldstone R/D High Speed Data Acquisition System

L. J. Deutsch, R. F. Jurgens, and S. S. Brokl
Communications Systems Research Section

A digital data acquisition system that meets the requirements of several users (initially the planetary radar program) is planned for general use at Deep Space Station 14 (DSS 14). The system, now partially complete, is controlled by a Digital Equipment Corporation (DEC) VAX 11/780 computer that is programmed in high level languages. A DEC Data Controller is included for moderate-speed data acquisition, low speed data display, and for a digital interface to special user-provided devices. The high-speed data acquisition is performed in devices that are being designed and built at JPL. Analog IF signals are converted to a digitized 50 MHz real signal. This signal is filtered and mixed digitally to baseband after which its phase code (a PN sequence in the case of planetary radar) is removed. It may then be accumulated (or averaged) and fed into the VAX through an FPS 5210 array processor. Further data processing before entering the VAX is thus possible (computation and accumulation of the power spectra, for example). The system will be located in the research and development pedestal at DSS 14 so easy access by researchers in radio astronomy as well as telemetry processing and antenna arraying is anticipated.

I. Overview

A High Speed Data Acquisition System is being constructed for research and development (R/D) use in the Deep Space Network (DSN). This system will, when completed, be a subsystem of the DSN Ground-Based Planetary Radar System. Its primary purpose will be to collect the 15 MHz bandwidth RF signal from the DSN Radar Receivers and perform all necessary data rate compression (filtering, phase correction, and correlating) to get radar data on a general purpose computer for further processing. This function is currently performed by an array of hardware devices that are, for the most part, inadequate for the type of high bandwidth radar experiments that will be performed over the next decade. A good idea of the capabilities of the current system can be seen in Refs. 1 and 2. This article provides a system level understanding of the new system.

The High Speed Data Acquisition System is a collection of advanced digital signal processing devices that are connected to a host minicomputer. The system will be configured to operate in a wide range of modes, allowing a user to interconnect these devices and control them completely under software control. The set of processing devices that will be available to users enables many types of radar experiments as well as pulsar observations, general radio astronomy, and some telemetry demodulation and decoding capabilities.

A Digital Equipment Corporation (DEC) VAX 11/780 computer is the heart of the system. The standard DEC operating system (VMS) will be used to run the machine while all of the general purpose peripheral devices will be connected to the VAX through conventional VMS driver routines. Consequently, nearly all data acquisition programs (i.e., software for the control of experiments) may be written in high level languages such as FORTRAN or Pascal. This design philosophy provides a means for low cost data collection and ease of documentation. It will also enable more people to operate the system with only a small amount of specialized training.

The configuration of the High Speed Data Acquisition System at completion is shown in Fig. 1. The top part of the figure represents the VAX computer and its peripherals while the bottom part shows the special purpose devices that will be built.

The VAX 11/780 computer is also configured with all the necessary peripherals for program development and testing. In addition there are commercial peripherals that are essential to the radar program, including a high resolution color video monitor for real time image display and a high resolution graphics printer for data monitoring during experiments. Data will be collected and saved on a 500 Mbyte Winchester disk or on two high speed magnetic tape drives. The storage capability on this system will be nearly a factor of ten better than that of the current radar system.

In order to process the high rate data imposed by anticipated signal bandwidths, special purpose devices are also required. These high speed digital signal processing devices will be connected to the VAX through a Floating Point Systems (FPS) array processor. Subsequently, the data may be further reduced and processed before being transferred to the VAX. This will relieve the typical throughput problems that one encounters when using a host computer to collect and process real time data.

In addition, there will be low rate data input and output available on the VAX by means of a set of standard DEC analog-to-digital and digital-to-analog converters. Data collected in this manner by the VAX can be transferred to the array processor for processing and then back to the VAX for handling or storage. The digital-to-analog converters can be used to drive instruments or real time displays in order to enhance certain experiments.

The radar correlator assembly consists of a block of sixteen correlator modules that will run at a clock rate of 10 MHz. Each module will be capable of performing 256 delay lags in a one by four bit correlation function in a continuous stream. Each individual module will be able to operate in a cross- or auto-correlation mode. The selection of input signals, reference signals, and interconnection of the modules will be under complete control by the host VAX. This means that a user can reconfigure the correlator assembly to suit the needs of each individual experiment, or change their function during an experiment in a controlled manner, all from software written in a high level language.

The radar system typically uses up to three complex input channels. However, a fourth complex channel is included in the new system. This extra channel provides for expansion as well as a hardware backup.

Since some radar experiments will require pulsed or spread spectrum radar as well as CW radar, pseudonoise (PN) sequence generators must also be part of the High Speed Data Acquisition System. One generator is required for each channel. The output of these generators will be supplied to the correlator assembly, the IF processing devices, and the transmitter driver. These PN generators are programmable (from the VAX) so that a large set of PN codes of various lengths can be used for different experiments.

A unique feature of the new system is the IF processor hardware. Input signals are sampled directly at the output of the IF amplifier which is centered at 7.5 MHz. With this center frequency, a channel bandwidth of 15 MHz is available. Thus, a sample rate of at least 30 MHz will be needed to handle the information band. A sampling rate of about 50 MHz will be used to digitize the IF signal as the input to the system. All signal conditioning and processing following this will thus be accomplished digitally. The conditioning processes will consist of antidispersion filters, baseband mixers, and baud integrators. The resulting baseband signal will then be fed to a set of signal multiplexers that will be used to distribute the various signals to the correlator assembly.

A wide range of preprocessing can be achieved, ranges from none to the full capability of filtering, correlating, and transforming. For this reason, it is better to view the High Speed Data Acquisition System as a four channel digital receiver system. Among the nonradar applications that this system will be capable of supporting are general radio astronomy, SETI (Search for Extra-Terrestrial Intelligence), and telemetry reception. These applications are addressed in more detail in the following section.

II. Applications

The primary mode of operation of the High Speed Data Acquisition system is the collection and processing of radar data. Since the system was designed with this purpose in mind, it will do this job particularly well. The system will support CW, pulsed, and spread spectrum radar experiments, and thus it will support experiments targeted at the outer planets as well as the terrestrial planets. A number of radar experiments anticipated over the next five years are listed below along with their system requirements. Figure 2 shows a time table of the various experiments.

Mars. This will be the first experiment that will use the High Speed Data Acquisition System. The observations are planned for June and July of 1984. Only one station (DSS 14) will be used. This experiment does not require the IF processor portion of the system. It will require at least one quarter of the correlator assembly, a PN sequence generator, and all of the computing facility. In addition, a baseband mixer and sampler assembly must be fabricated (mostly out of existing parts) to interface the system to the IF output of the radar receiver.

Mercury. Mercury can be observed several times each year. This work will require either a dual polarization receiver or the use of two stations operating at different polarizations. In either case, at least two channels of the IF processor must be in place. Mercury will be observed with both CW and spread spectrum radar.

Venus. Venus will be observed in the first six months of 1985. These observations require three stations. The entire IF processor must be working by this time. Also, the full capability of the correlator assembly will be needed.

Asteroids. These observations will be made on a semiregular basis. CW radar with frequency hopping and a dual-polarized receiver will be used.

Comets. These observations are similar to those of asteroids. Of particular note are the proposed observations of Halley in early 1986.

Moons of Jupiter. These observations will resume in April of 1985. They will also use CW radar with frequency hopping and dual-polarization.

Rings of Saturn. The first observation will be in May of 1985. CW radar with frequency hopping and dual-polarized signals will be used.

The High Speed Data Acquisition System will also be a useful tool in general Radio astronomy. It could be used to measure Stokes parameters (polarization parameters) and average the results in real time. It will also be possible to run the system as a real time interferometer by using the multiple channels in the IF processor.

The flexibility that is being designed into the High Speed Data Acquisition System will make it a useful tool in DSN receiver development. It will be the first system at a DSN site to have a fully digital IF processor. Since all aspects of this IF processor are to be programmable from the VAX computer, it can be used to experiment with new types of mixing and filtering techniques for telemetry and radiometric reception. It will be possible to control the loop filters dynamically during reception to provide improved receiver performance. The four channels in the IF processor can be used to perform real-time arraying of up to four antennas as well as carrier arraying for telemetry reception. The correlator assembly could also be used to demodulate a subcarrier and track data in a telemetry signal. Telemetry decoding could then be done on the VAX if the data rates are low enough. If the system were to be used for extensive high rate telemetry experiments, then a Viterbi decoder (MCD) could be attached to the VAX as another high speed peripheral.

The extreme flexibility in the High Speed Data Acquisition System and the ability to control this flexibility from simple high level programming languages is certain to make the system a major research tool for many projects well into the future.

III. Computer Facility

The High Speed Data Acquisition System is based on a Digital Equipment Corporation VAX 11/780 computer. This computer was chosen to meet the need for high-performance real time data collection, data processing, and program development. This machine, and all its associated peripherals and hardware add-ons, will be located at the Goldstone facility in the pedestal of DSS-14.

The VAX computer uses a 32-bit architecture that is based on the PDP-11 family of 16-bit minicomputers. Although the VAX uses addressing modes and stack structures that are similar to those on PDP machines, the VAX provides 32-bit addressing. This is enough to access a four gigabyte program address space. The 32-bit data paths in the VAX also offer increased processing speed and accuracy over the PDPs. The VAX processor hardware and instruction set were specifically designed to support high-level programming and still produce efficient machine code. The VAX also has a floating point processor that augments the primary processor by performing floating point operations with a substantial performance improvement. There is also a high speed memory cache that provides a large reduction in memory access time. The VAX main memory uses error correcting codes to correct single bit errors and detect double bit errors in each byte.

The VAX operating system, VAX/VMS (or simply VMS) is a multi-user virtual memory system that supports an easy-to-use interactive command interface. The virtual memory feature allows the execution of programs that can be larger than the physical memory allocated to them. Virtual memory paging (the operation that swaps memory and disk space to allow this feature) is handled automatically by VMS. Consequently, the programmer need not be concerned with the working details of memory management. Memory management facilities can be controlled by the user in the event that such control is necessary to optimize the real-time performance of a particular program.

There are two hardware buses on the VAX backplane to which peripherals may be attached. One of these is the MASS Bus. It is capable of high throughput communications. The second, the UNIBUS, provides a standard interface that can be used with a large set of peripherals. Figure 3 shows the configuration of the High Speed Data Acquisition VAX. The blocks that have shaded corners represent equipment that is already installed and working in a JPL Laboratory. The peripherals that will be attached to the VAX are listed below (but do not include the special purpose hardware devices that will be described in the following sections):

2 RK07 28 Megabyte Disk Drives. These drives use removable hard disk cartridges. These are useful for storing the operating system and most user software. Since they are removable, they provide a fast method for transporting software or performing backups.

1 RP07 500 Megabyte Disk Drive. This device, which is to be attached to the MASS bus, will be a principal storage area for collected data.

2 TU78 Magnetic Tape Drives. These are high speed (1600 or 6250 bpi at 125 ips) tape drives that can be used in real time data acquisition or high speed data transfers.

1 DR780 High Speed 32-bit Parallel Interface. This device is a 32-bit parallel interface that is capable of transferring data at rates up to two megabytes per second. The DR780 uses separate data paths for control information and data. The control path is 8-bit and bidirectional. The data path provides a synchronous 32-bit transfer which may be clocked by either an LPA11-K clock or by a clock provided by an external device.

1 LPA11-K Analog/Digital Data Controller. This device is an intelligent controller that transfers data between the VAX computer and a set of interface modules. These modules include analog-to-digital and digital-to-analog converters and a high speed 16-bit parallel interface (see Section IV). The LPA11-K subsystem uses silo buffers and direct-memory-access data transfers to allow real-time data acquisition with a very limited interrupt load and very limited burden in the VAX itself. The LPA11-K is implemented using two microprocessors, and it contains its own programmable real-time clock. The VAX can initiate a chain of data transfers through the LPA11-K with only a single I/O request. This silo structure allows data to be transferred continuously regardless of the status of the UNIBUS or system software latencies. Data rates up to 15,000 samples per second can be handled by the LPA11-K. There is also a mode where the LPA11-K services only the analog-to-digital converter and can handle 150,000 samples per second.

6 DR11-C 16-bit Parallel Interfaces. These will be used to provide control to the various special purpose hardware that will be described in Section IV.

2 VT100 CRT Terminals. These will be the primary user interface with the VAX for both program development and experiment control.

1 AED512 Color Graphics Terminal. This will be used as a real time high resolution display for monitoring experiments and reviewing collected data. The AED512 provides a 512 X 512 pixel display with 8-bit resolution in each of the three primary colors for a total for 24 bits for each pixel.

1 LA120 Operator Console. This device will be used to monitor the VAX itself and keep hardcopy logs of system performance. Hardware problems will be reported to the console so that the operator may act upon them when necessary.

1 LXY12 Printer/Plotter. This device will be the primary hardcopy output generator. It will be used to examine quick-look data analysis and to aid in software development.

1 FPS-5210 Array Processor. This device is equipped with a programmable I/O processor (GPIOP) that can be used to collect data from an external device and pass it through the array processor independently of the VAX. Maximum data rate through the GPIOP is three million 38-bit words per second. The FPS 5210 is rated at 30 million floating point operations per second. It will be used to format and process incoming data before it is sent to the VAX.

1 DEC PDS Power Distribution System. The PDS will provide filtered and regulated power to all the VAX peripherals.

The VAX provides the High Speed Data Acquisition with the required flexibility and expandability that are needed to support experimental work at Goldstone well into the future.

IV. General Purpose Acquisition Hardware

The High Speed Data Acquisition System will have a set of peripheral devices that perform conventional data acquisition functions. These functions include digital-to-analog conversion, analog-to-digital conversion, and high throughput digital-to-digital interfaces. These can be used to interface the system with existing data acquisition hardware, laboratory instruments, or specialized display devices. They will also be instrumental in the testing of the system itself.

The general purpose acquisition hardware will reside on its own dedicated bus so that a continuous data flow may be achieved regardless of the status of the buses on the VAX itself. This bus will be attached to the VAX through the LPA11-K Analog/Digital Data Controller that was described in Section III. Figure 4 shows the configuration of this part of the system. The shaded boxes, once again, represent hardware that is already installed in the laboratory.

There will be four modules attached to this LPA11-K bus that will comprise the general purpose acquisition hardware. These are described below:

AA11-K 4-Channel D/A and Display Control. This device provides four independently buffered digital-to-analog converters. Each of these has 12-bit accuracy and a slew rate of 5V/s. In addition the AA11-K includes a display control driver that supports a 4096 × 4096 dot array display. This can be used to drive a high-resolution graphics CRT or an oscilloscope.

AD11-K Analog to Digital Converter. The AA11-K contains a high speed 12-bit analog-to-digital converter and an associated multiplexer. The multiplexer can be used to support 8 full channels or 16 single-ended channels of input. The timing for the conversions can be provided in one of three ways: under software control, from a real-time clock, or from an external input. A multichannel throughput rate of 50 KHz will be possible.

DR11-K Interface. This is general purpose 16-bit parallel interface for interconnection with other digital devices.

KW11-K Dual Programmable Real Time Clock. This module provides two real-time clocks that will be used to control data flow in and out of the LPA11 subsystem and provide accurate time tagging. Each clock is independently programmable and may operate at several different clock frequencies, as may be required for various applications.

These devices, together with the LPA11-K controller, will constitute a subsystem of the High Speed Data Acquisition System that will provide a general compatibility with other data acquisition and data processing hardware without disturbing the other portions of the system. The inclusion of this subsystem will make it easier to append hardware devices to the system when such devices are needed without having to install these permanently.

V. High Speed Data Acquisition Hardware

In addition to the computing facility and the general purpose data acquisition hardware, the High Speed Data Acquisition System will include special purpose hardware that will enable it to collect and process data directly from the IF portion of the radar receiver. This same hardware could be used to interface with other receivers as well. Figure 5 shows the special purpose hardware and its relationship to the computing facility. Most of this special purpose hardware is not available commercially and hence it must be designed and built in-house. The hardware with shaded boxes in the figure has been fabricated and tested.

The high speed data acquisition hardware comprises three subsystems. These are the IF processor subsystem, the correlator assembly, and the PDPG coders (PN sequence generators). Figure 5 shows these subsystems and their relationship to the High Speed Data Acquisition System computing facility.

The IF processor will serve as the principal path for input data in the completed system. It will be designed to accept, as input, an analog IF signal with a carrier of 7.5 MHz and a two sided bandwidth of 15 MHz. Thus the analog-to-digital converters (A/Ds) must sample at no less than 30 MHz and will do so with four-bit accuracy. However, in order to assure that there is a sufficient rejection of frequencies outside of this bandwidth, the converters will actually be capable of a 50 MHz sampling rate.

The outputs from the converters will be passed through Anti-Dispersion Filters (ADFs) before mixing so that phase corrections can be made on a controlled basis. These filters will be programmable from the VAX.

The Digital Complex Mixers (DCMs) will be high speed (50 MHz) digital multipliers, which will heterodyne the IF signal down to baseband. The reference frequency for the conversion can come from the VAX (predictions stored in memory and sent to a PLO) or from feedback within the system.

The signal will then be passed through Complex Baud Filters (CBFs). These devices will serve two purposes. They will act as loop filters that can be controlled by the VAX. In addition, they will smooth the signal so that it may be sampled at a lower rate for subsequent processing.

The correlator assembly consists of a baseband signal multiplexer, correlator-accumulator modules, and a device for data formatting and interfacing to the array processor.

The baseband multiplexer will allow the baseband signals from the four complex channels of the IF processor to be routed to any combination of correlator-accumulator modules. This function will be under the control of the VAX. In this way, the correlator assembly can be configured many ways, depending on the needs of a particular experiment. The correlators can have a variable number of lags (selectable under computer control) and operate as cross- or auto-correlators.

Each correlator-accumulator module will include a complete 4-by-1 256 lag correlator and a pipelined adder that will produce a 12-bit output. These will operate at a clock rate of 10 MHz. The accumulators will be capable of summing the correlator outputs for up to one full second. The output of each module will be a 28-bit word.

The Interface and Data Formatting will collect the output from the correlator-accumulator modules and feed that data through the General-Purpose Input/Output Port (GPIOP) of the array processor. This collecting function will also be under the direct control of the VAX.

The PN sequences that will be used for spread-spectrum radar work will be supplied by the polynomial driven PN Generators (PDPGs). Each of these will run at a 10 MHz chip rate that will be synchronized with an accurate reference clock. The PDPGs will also provide time domain compensation for the Doppler effects caused by the relative motion of the observer and the target. The PDPGs will be controlled by the VAX and will be capable of generating a large range of PN sequences from a single pulse to a sequence $2^{24} - 1$ bits long.

This special purpose hardware will compose the first high-speed digital IF receiver to be installed at Goldstone. It will be capable of serving as a test bed for digital receiver work as well as radar and radio astronomy work.

VI. Interfacing to the Remainder of the Radar System

In order to perform the experiments that were described in Section II, the High Speed Data Acquisition System must interface with other DSN systems, including transmitters, receivers, communication links, and frequency reference generators. These systems are also undergoing fabrication or upgrading. It is essential that they are operating at the time that they are needed for the various experiments. We describe here the major requirements that the High Speed Data Acquisition System will put on these other systems:

(a) Radar experiments must be performed with transmitters and receivers that operate on the same frequencies. Both S- and X-band signals will be required.

(b) The bandwidth of the transmitter, receiver, and antenna systems must be adequate to pass a 2 MHz PN phase code without a serious degradation of shape for any length $2^N - 1$ where $1 \leqslant N \leqslant 30$.

(c) The X-band bandwidth must also support the Rings of Saturn experiment. At least 10 MHz must be passed for this.

The receiver must provide both orthogonal-circular polarizations simultaneously at X-band. Two identical channels are required, each having a 10 MHz bandwidth centered on a frequency of 8495 MHz. A system noise temperature of less than 20 K is preferable for this work. A similar S-band capability should also exist although it is not mandatory for the proposed radar experiments.

Some experiments require that stations other than DSS-14 be used during an observation. To this end, interstation links must exist between DSS-14 and both DSS-12 and DSS-13. These links must be capable of supporting the PN signal described above.

Accurate timing signals must be provided to the acquisition system. Most of the work can be supported by rubidium clocks. However, any observations that will use more than a single station will require hydrogen clocks.

VII. Conclusion

The High Speed Data Acquisition System will serve the needs of the DSN's Planetary Radar Program at the Goldstone facility over the next decade. The system will be useful in other areas of research including radio astronomy, SETI, and real time antenna arraying. The ability to control the system in a high level software language will enable it to be used by a larger group of people than the current system, and this flexibility will also facilitate its use as a tool for advanced receiver and demodulator development.

References

1. G. S. Downs and P. E. Reichley, "Radar Ranging of the Planet Mars 8495 MHz," *Deep Space Network Progress Report 42-29*, 1975, pp. 95-106.
2. G. S. Downs and P. E. Reichley, "Instrumental Polarization of the Goldstone 64-m Antenna Systems at 2388 MHz," *Deep Space Network Progress Report 42-27*, 1975, pp. 112-116.
3. S. S. Brokl, "Polynomial Driven Time Base and PN Generator," *TDA Progress Report 42-75*, 1983, pp. 84-90.

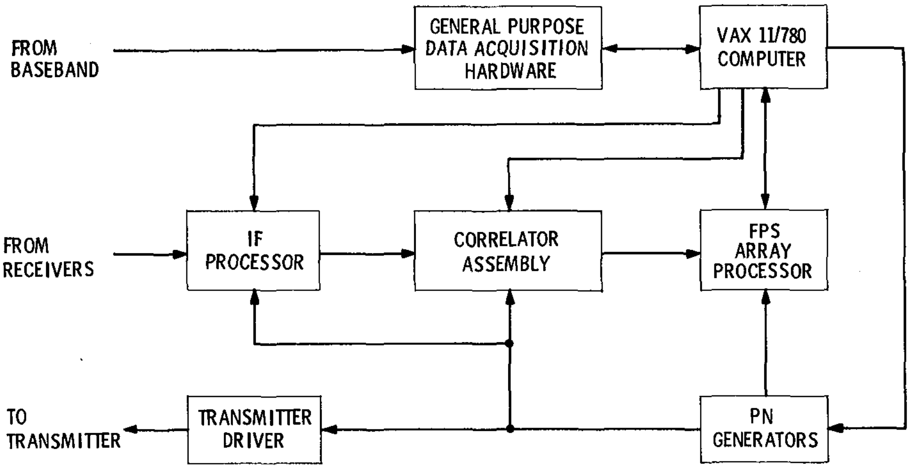


Fig. 1. High Speed Data Acquisition System overall block diagram

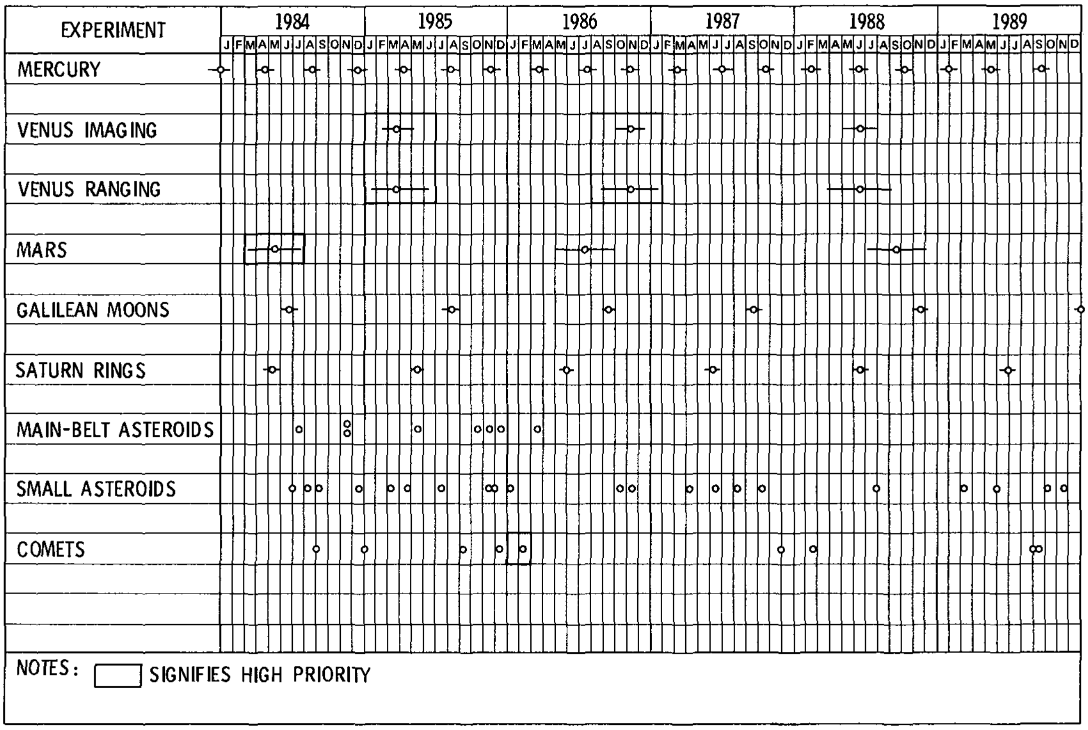


Fig. 2. Time table of anticipated radar experiments

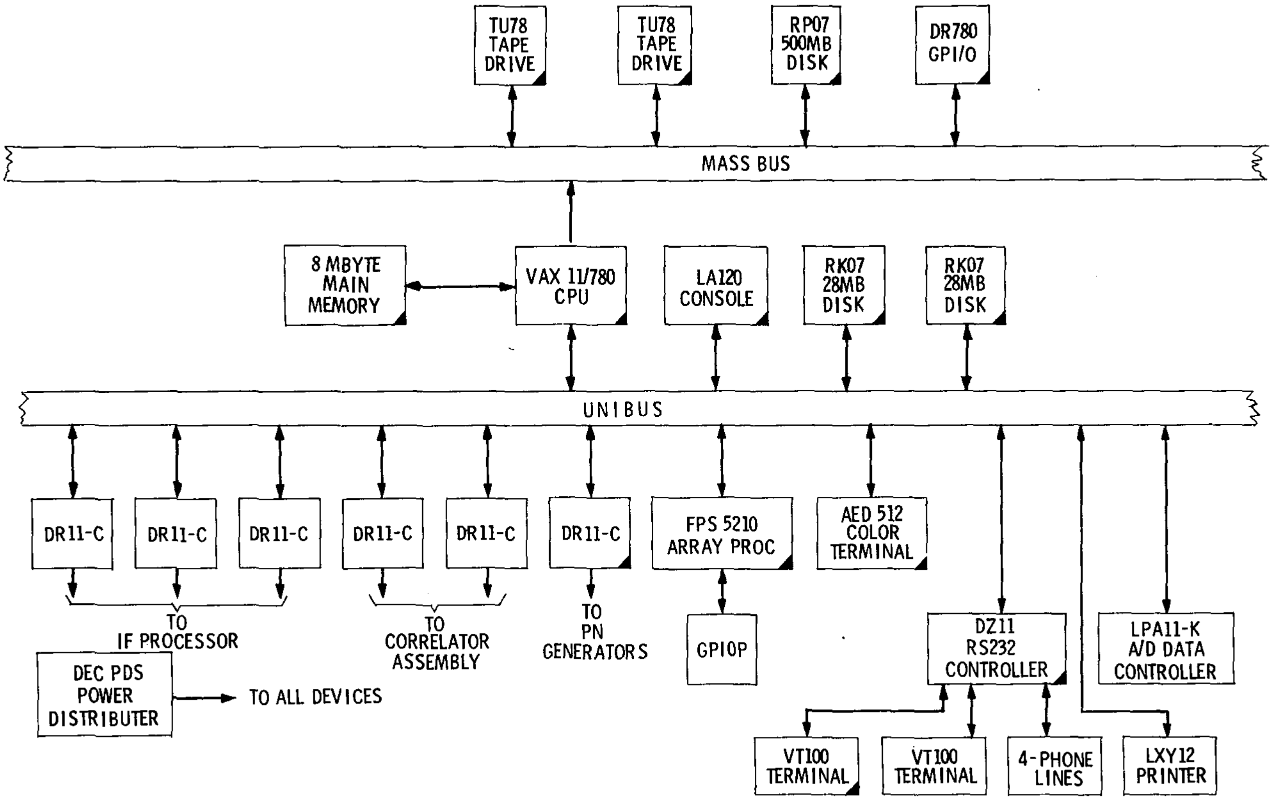


Fig. 3. The VAX 11/780 configuration for the HSDAS. Shaded corners represent equipment that is installed and working at JPL.

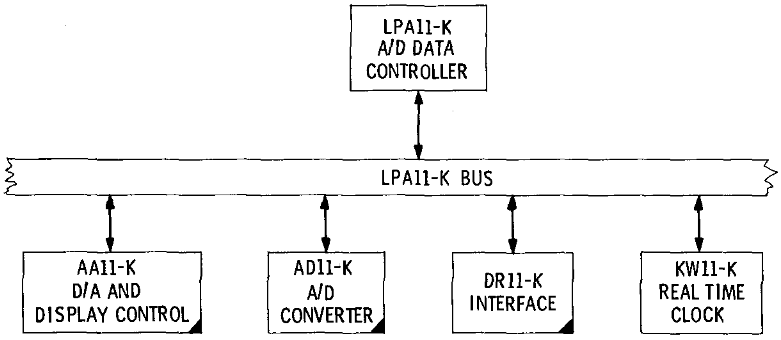


Fig. 4. General purpose data acquisition hardware for the HSDAS. Shaded corners represent hardware that is already installed.

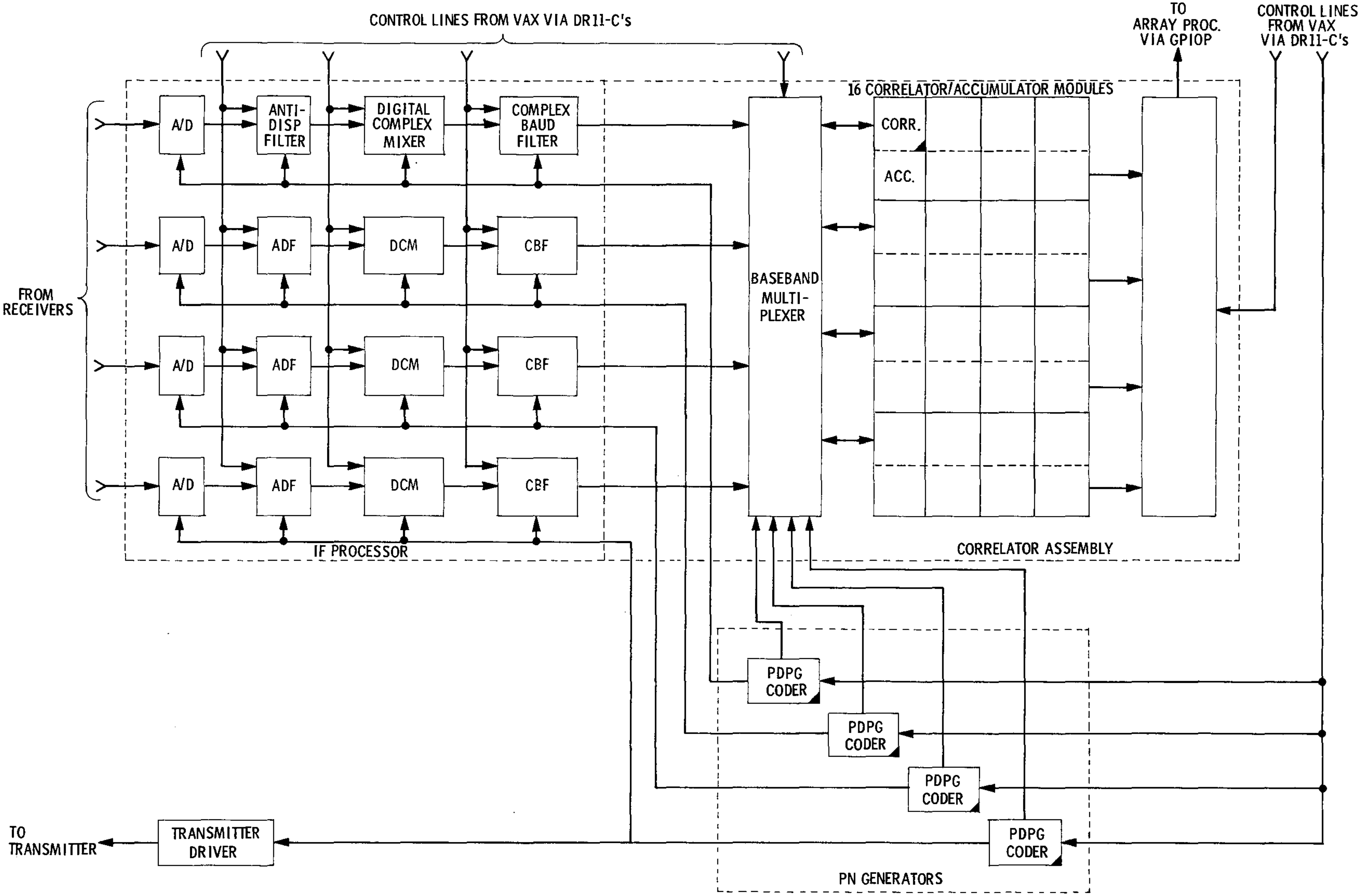


Fig. 5. Special purpose data acquisition hardware. Hardware with shaded boxes has been fabricated and tested.

Demodulator and Accumulator for the High-Speed Data Acquisition System

S. S. Brokl
Communications Systems Research Section

In support of nonflight project data acquisition, a new High-Speed Data Acquisition System was designed. The demodulator and accumulator modules are at the heart of this new system and are described in this document.

I. Introduction

The High-Speed Data Acquisition System (HSDAS) is currently being implemented principally to support the DSN Planetary Radar Program (Ref. 1). In the HSDAS, the Polynomial-Driven PN Generators (Ref. 2) will be used to bounce a coded waveform off the radar target. The demodulator and accumulator subsystems are to be used to detect the reflected signals and produce ranging information for further processing.

Each demodulator consists of 16 TRW TCC1023J digital correlator chips configured as 4 by 1 bit auto or cross correlators, with 256 lags. Included on the module is all of the logic and adders required to represent the 256 lag data as a 12 bit offset binary number.

The accumulator module follows the demodulator and allows accumulations of 12 bit offset binary numbers from each lag to a maximum 28 bit number. Each accumulator module supports two demodulators. For the final system there will be 16 demodulator and 8 accumulator modules.

II. The Demodulator

Figure 1 is a block diagram for the demodulator. There are 16 TRW TCC1023J correlators which are configured as 4X1 auto or cross correlators. The autocorrelation function is:

$$C_{(n)} = \sum_{\ell=0}^{256} C_{\ell} \left[G_{\ell+n}\right]$$

where G_{ℓ} is a one bit representation of the original data, $G_{\ell+n}$ is a four bit representation of the original data time-shifted, and ℓ is the lag number.

The cross-correlation function is:

$$C_{(n)} = \sum_{\ell=0}^{256} G_{\ell} \left[H_{\ell+n}\right]$$

where G_{ℓ} is a binary signal, $H_{\ell+n}$ is a four bit representation of the returned signal time-shifted, and ℓ is the lag number.

The normal modes of operation for the demodulator in the radar acquisition system will be as a cross correlator, when doing delay Doppler gating using the Polynomial Driven PN Generator, or as an autocorrelator when doing continuous wave (CW) detection or radio astronomy.

It will be possible to configure several of the modules within the system to run as a 512 lag or 1024 lag correlator, with both real and imaginary channels. It may even be possible to run 8 by 1 bit correlations.

The correlator chips accomplish the multiply with exclusive-nor logic (Ref. 3). The truth table is shown below.

A	B	$\overline{A\ B}$
0	0	1
0	1	0
1	0	0
1	1	1

This can be thought of as a multiplier if one replaces 0 by -1 and 1 by $+1$. This implies that the A/D converter data input to the demodulator should be expressed in offset binary. A conversion table for a four bit 2's complement to offset binary number is shown in Table 1. As can be seen, the transformation only requires toggling the sign bit.

In the demodulator Fig. 1 there are 16 correlator chips broken into 4 groups of 4 chips each. The adder tree that follows sums the results of the correlators into a 12 bit number representing a running total of correlations for four bits over 256 lags.

The correlators also have a mask input so that any number of bits may be masked off and will not be used as part of the total output calculation. This is extremely useful when running short codes since it prevents the multiple copies of unused lags from being summed.

The demodulator uses one master clock which can run as high as 20 MHz. The code, mask, and signal have their own enable inputs to the module. This allows the clock to bring the data as required.

The correlator chips contain two shift-register chains with one isolated from the exclusive-nor logic by a register which can be parallel loaded with the 64 bit serial bit stream in one of the shift-register chains. The signal LDR into the demodulator does this function for 256 lags. This allows a new 256 bit code string to be serially loaded while correlating the first stream, which is being held in the registers.

Figure 2 is a photograph of the demodulator module. The modules for the radar system have been built and tested. They have also been operated successfully in a system test with the Polynomial Driven PN Generators.

III. The Accumulator

Figure 3 is a block diagram of one half of the accumulator module. Each module contains two complete 28-bit wide accumulators. Each accumulator has two buffer memories (double buffered), so that one memory may be unloaded by the computer while the other memory is accumulating data. The accumulation function is:

$$A_{(\ell)} = \sum_{m=1}^{M} C_{(\ell,m)}$$

where $\ell = 1{,}256$ is the lag number and M is the number of accumulations. M is programmable from 1 to 2^{16}.

The purpose of the accumulator is to allow a correlated signal to be integrated for a period m times longer than the storage length of the correlator.

The control of the accumulator is derived in the computer interface and formating module which is being designed. The accumulator is designed with Fairchild FAST tm logic and 30 ns access time memories. The accumulator will run with a maximum clock rate of 10 MHz.

Figure 4 is a photograph of the accumulator module. The design and construction are complete.

IV. Conclusion

The demodulator and accumulator modules perform the detection and prefiltering of data which will be further processed by a Floating Point Systems FPS 5210 Array Processor. The array processor is also used to form the complete sums when multiple units are used to extend the number of lags or the number of bits in the product table. The High-Speed Data Acquisition System was designed for High Resolution Radar Astronomy. The system has applications in astronomy and possible application for the Search for Extraterrestial Intelligence (SETI) and perhaps RFI monitoring.

References

1. L. J. Deutsch, R. F. Jurgens, S. S. Brokl, "The Goldstone R/D High Speed Data Acquisition System," *Telecommunications and Data Acquisition Progress Report 42-77*, January-March 1984, Jet Propulsion Laboratory, Pasadena, CA.
2. S. S. Brokl, "Polynomial Driven Time Base and PN Generator," *Telecommunications and Data Acquisition Progress Report 42-75*, July-September 1983, Jet Propulsion Laboratory, Pasadena, CA, pp. 84-90.
3. John Eldon, "Correlation . . . A Powerful Technique for Digital Signal Processing," *LSI Publication TP17B-4/81*, TRW LSI Products, TRW Inc., 1981.

Table 1. Conversion table for a four bit 2's complement to offset binary number

Range	2's Complement	Offset Binary
+MAX	0 1 1 1	1 1 1 1
	0 1 1 0	1 1 1 0
	0 1 0 1	1 1 0 1
	0 1 0 0	1 1 0 1
	0 0 1 1	1 0 1 1
	0 0 1 0	1 0 1 0
	0 0 0 1	1 0 0 1
0	0 0 0 0	1 0 0 0
	1 1 1 1	0 1 1 1
	1 1 1 0	0 1 1 0
	1 1 0 0	0 1 0 0
	1 0 1 1	0 0 1 1
	1 0 1 0	0 0 1 0
	1 0 0 1	0 0 0 1
–MAX	1 0 0 0	0 0 0 0

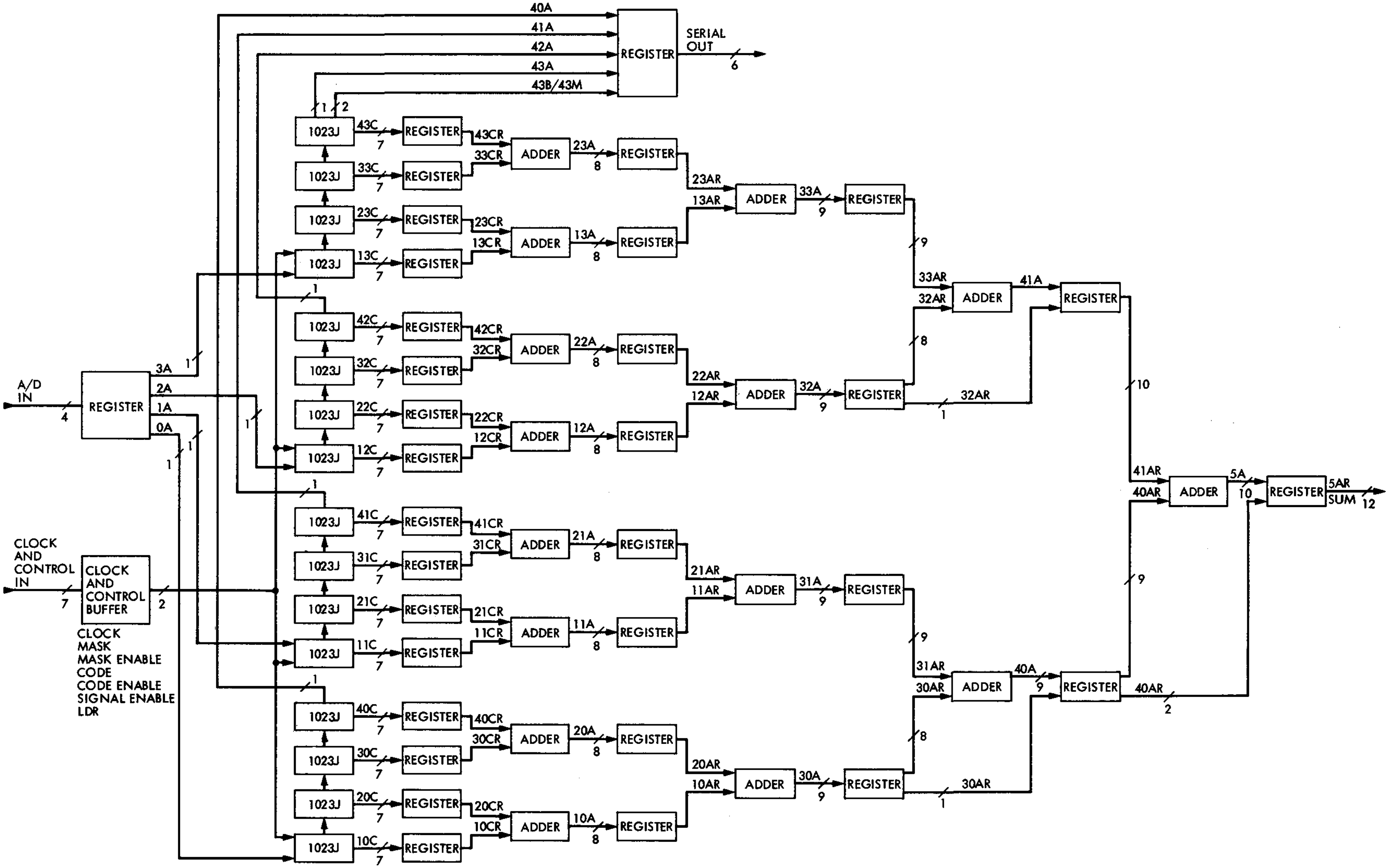


Fig. 1. Demodulator block diagram

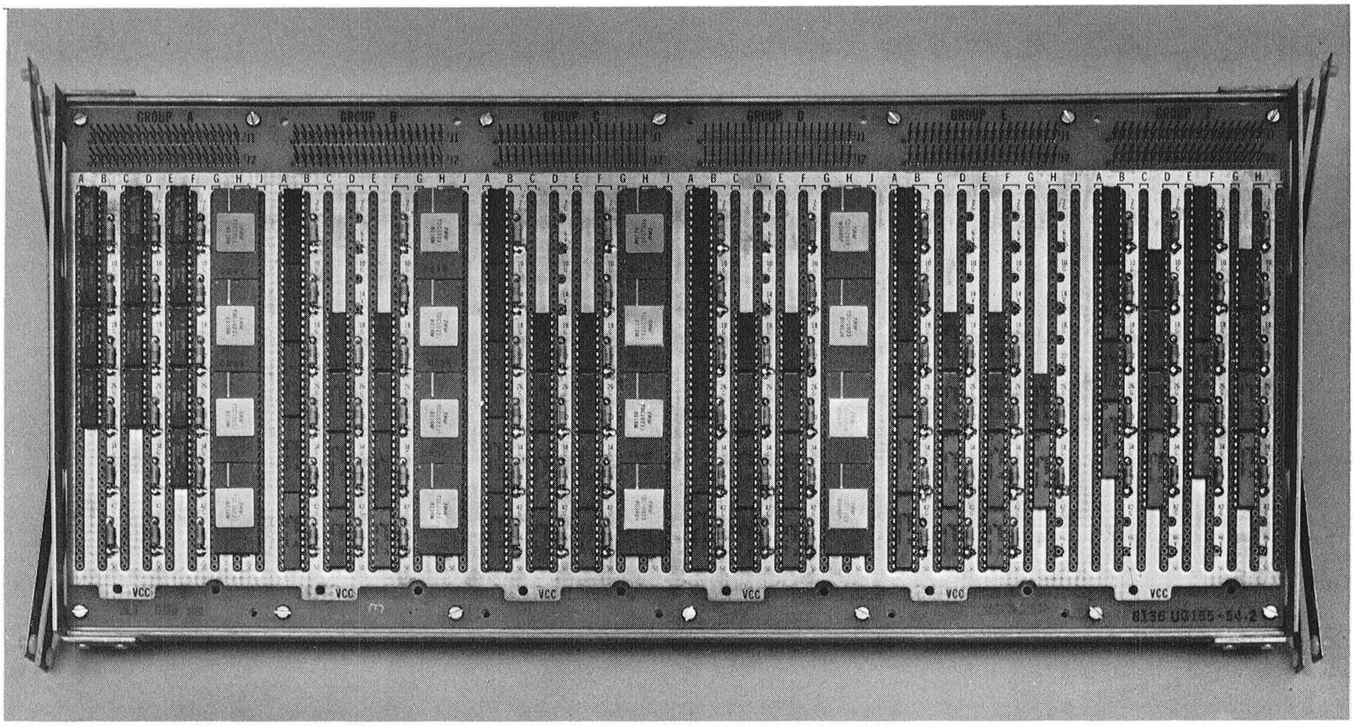
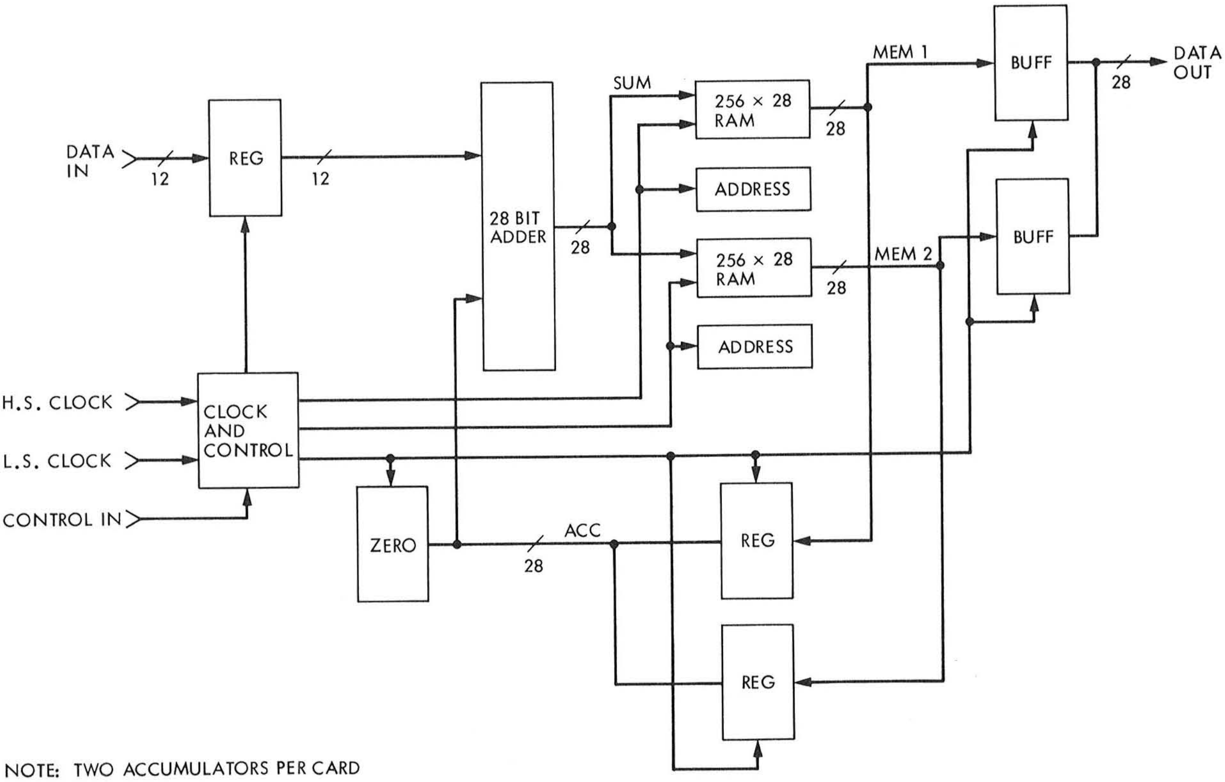


Fig. 2. Demodulator module



NOTE: TWO ACCUMULATORS PER CARD

Fig. 3. Accumulator block diagram

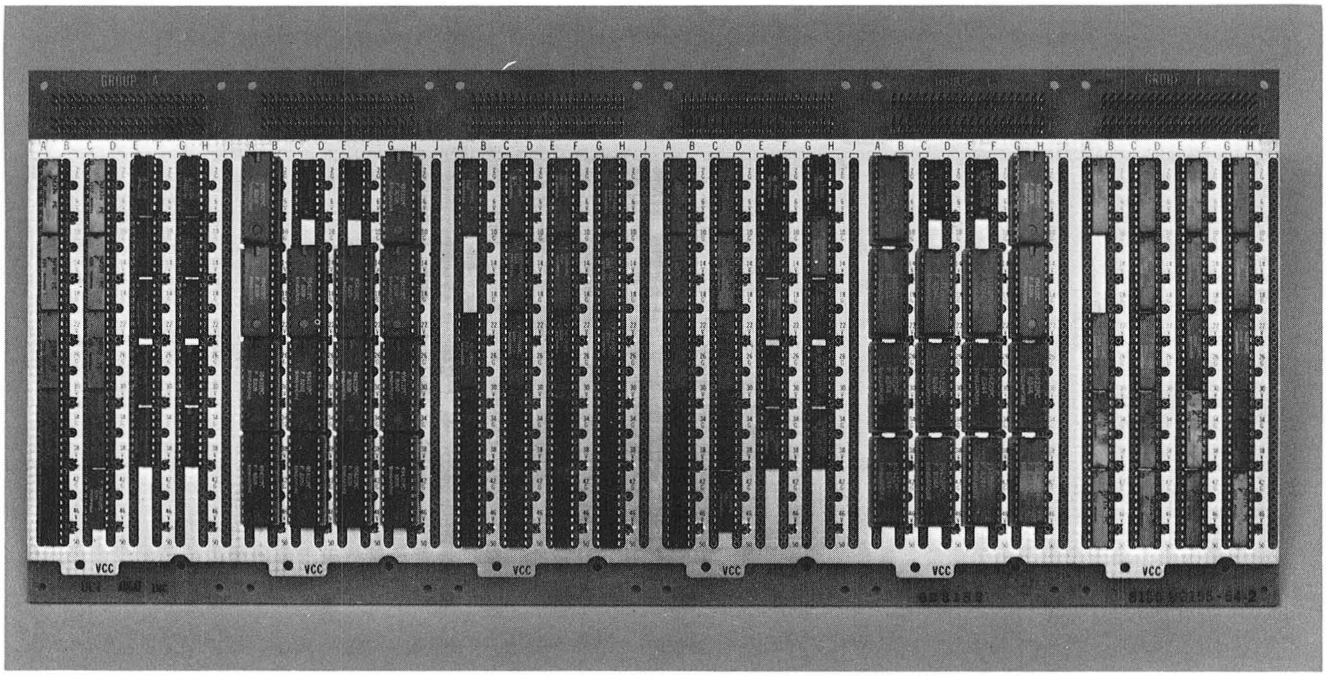


Fig. 4. Accumulator module

Effect of RF Filtering on the Performance of Uncoded PCM/PM Telemetry Channels

M. A. Koerner
Telecommunications Systems Section

This article describes a method for calculating the increase in received telemetry signal power required to compensate for the use of a radio frequency interference (RFI) filter in front of the DSN receiving system low-noise amplifier. The telemetry system for which the degradation is calculated is an uncoded PCM/PM system in which the NRZ data directly modulates the carrier at a modulation level which leaves a discrete carrier. A phase-locked loop in the RF receiver tracks the discrete carrier and coherently demodulates the PCM data. The RFI filter may be a series of Butterworth, Tchebychev, or Bessel low-pass, high-pass, band-pass, or band-reject filters, each with arbitrary bandwidth, number of poles, and, for band-pass or band-reject filters, resonant frequency. The only restriction is that the RFI filter must have only simple poles. Numerical results are presented for the RFI filters the DSN plans to place in front of the S-band FET, S-band maser, and X-band maser low noise amplifiers. The main conclusion is that the filters will produce negligible degradation at data rates below 4 Mbps.

I. Introduction

In conjunction with the Networks Consolidation Program (NCP), the DSN is developing band-pass filters to be placed in front of the S-band FET (field-effect transistor) low-noise amplifier (LNA) and the S-band and X-band maser LNAs. The purpose of these filters is to avoid LNA saturation by out-of-band radio frequency interference (RFI) signals. The DSN itself produces some of this interference, such as that from the DSN transmitters. The remainder comes from other systems operating in adjacent frequency bands. Reference 1 reports the results of a study to define the requirements for these filters.

The use of band-reject filters has also been considered. Such filters could be used to eliminate both in-band and out-of-band RFI signals that have fixed frequency and narrow bandwidth.

An unavoidable consequence of the use of these filters will be the degradation in telemetry performance caused by the telemetry signal distortion produced by these filters. The term "telemetry degradation" as used in this report is the increase in received signal power necessary to compensate for the addition of the RFI filter. Thus, the degradation will depend on the basis (i.e., bit error probability) on which telemetry performance is measured. The degradation will be greatest for very wide-band telemetry channels operating near the edge of the RFI filter pass-band. The objective of this analysis is (1) to develop a method of calculating this telemetry degradation for a variety of RFI filters, and (2) to apply this method to the RFI filters being developed by the DSN.

The telemetry system considered in this report is an uncoded PCM/PM system. A very simplified block diagram of this system is shown in Fig. 1. NRZ (non-return-to-zero) data directly phase modulates the RF carrier at a modulation level which leaves a discrete carrier component. At the DSN, a phase-locked loop in the RF receiver, which follows the RFI filter and LNA, tracks the received carrier and demodulates the uncoded NRZ PCM signal. The RF receiver is followed by an integrate-and-dump circuit, which integrates the RF receiver output over one bit time. A bit synchronizer establishes the timing for the integrate-and-dump circuit. At the end of the integration time for each bit, a binary decision is made as to whether a PCM "0" or "1" was transmitted, based on the polarity of the integrate-and-dump circuit output.

The primary alternatives to the telemetry system analyzed in this report are (1) an uncoded PCM/PM system where the carrier is completely suppressed and a Costas loop recovers the carrier reference, and (2) an uncoded PCM/PSK/PM system where the NRZ PCM data phase shift keys a subcarrier and the resulting signal phase modulates the carrier. While some work was done on both of these systems, it was decided to concentrate on the discrete carrier PCM/PM system because this is the system most likely to be used, at least initially, by the DSN for the very wideband telemetry channels for which significant degradation is expected from signal distortion in the RFI filter.

While this analysis considers only uncoded PCM/PM channels, it would be desirable to analyze systems such as Viterbi or Reed-Solomon/Viterbi channels, where the raw data are coded and the resulting coded symbol stream phase modulates the carrier. However, with the exception of Manchester coding, to which the results obtained herein can be easily extended, an analysis for coded channels, similar to that presented here for uncoded channels, does not appear feasible. Note, however, that the results presented here are equally valid for coded channels, as long as the channel symbol error probability is used as the measure of telemetry channel performance and the probability of symbol sequences in the coded PCM is the same as that in random data.

This analysis employs two important approximations. First, the analysis assumes that the RF receiver phase-locked loop tracks the received carrier with some constant phase error ϕ. The RF receiver phase-lock loop phase errors caused by thermal noise and the portion of the telemetry sidebands that fall within the bandwidth of the RF receiver phase-locked loop are neglected. Thus, any interaction between the signal distortion caused by the RFI filter and the time-varying RF receiver phase-locked loop phase error will not be uncovered by this analysis. Second, this analysis does not determine the actual displacement in bit synchronizer bit timing caused by the signal distortion. The analysis supplies a method of calculating the degradation as a function of the displacement in bit synchronizer timing. A first estimate of the displacement in bit synchronizer timing is made using the slope at zero frequency of the phase response of the RFI filter's baseband equivalent, and the telemetry degradation is calculated for this bit timing. This timing displacement will be that seen by very narrowband telemetry channels. Then, using this estimate of bit synchronizer timing displacement as an initial value, the bit synchronizer timing which minimizes the telemetry degradation and the degradation for this bit synchronizer timing is calculated. The actual bit synchronizer timing displacement may not be at the optimum value. Of course, it would always be possible to manually adjust bit synchronizer timing to the optimum value, but this may not be practical for normal telemetry operations.

II. Analysis

The objective of this analysis is to provide a method of computing the telemetry channel performance degradation caused by the RFI filter. Each of the subsections into which this section is divided provides the equations necessary for one step in the computational process. Because of the complexity of the problem, no attempt will be made to analytically combine the results presented in these subsections.

A. RFI Filter Poles and Residues

A basic assumption of this analysis is that the RFI filter has only simple poles. The transfer function of such a filter can be expressed in the form

$$H(s) = Q_0 + \sum_{k=1}^{N} \frac{Q_k}{s - P_k} \qquad (1)$$

where Q_0 is the filter response at infinite frequency, P_k is the kth pole of the filter, and Q_k is the residue at that pole. The first step in calculating the telemetry degradation caused by the RFI filter is to calculate the response at infinity and the poles and residues for the RFI filter.

This article considers an RFI filter consisting of an arbitrary number of cascaded elements. Since the RFI filter has only simple poles, each of its elements must have only simple poles and therefore a transfer function of the form shown in Eq. (1). Each of these elements can be a Butterworth, Tchebychev, or Bessel low-pass filter or the high-pass, band-pass, or band-reject filters which can be synthesized from these low-pass filters using simple frequency transformations. Thus, this analysis provides the capability for assessing the telemetry degradation produced by using both a band-pass RFI filter for rejection of out-of-band RFI and one or more band-reject filters for eliminating in-band RFI.

The first step in calculating the RFI filter response at infinity, its poles, and the residues at these poles, is to calculate these parameters for the low-pass equivalent of each RFI filter element. The response at infinity and the poles and residues for RFI filter elements that are high-pass, band-pass, or band-reject filters will be calculated from the response at infinity and the poles and residues of these low-pass equivalents. The low-pass equivalent's frequency response will be adjusted to have unit response at zero frequency and either amplitude response A at some frequency $f_A = (2\pi)^{-1}\,\omega_A$ or noise bandwidth B_N and will have zero response at infinite frequency. The transfer functions of the Butterworth, Tchebychev, and Bessel low-pass filters considered in this article have no zeros. Therefore, if $P_0, \ldots, P_N$ are the poles of the low-pass equivalent, the residue at the kth pole will be

$$Q_k = -P_k \prod_{\substack{\ell=1 \\ \ell \neq k}}^{N} \left(1 - \frac{P_k}{P_\ell}\right)^{-1} \tag{2}$$

1. Butterworth low-pass filters. The poles of a Butterworth low-pass filter with N poles will be

$$P_k = -\,\omega_B \sin\left[(2k-1)\,\frac{\pi}{2N}\right] - i\omega_B \cos\left[(2k-1)\,\frac{\pi}{2N}\right], \qquad 1 \leqslant k \leqslant N \tag{3}$$

where $f_B = (2\pi)^{-1}\omega_B$ is the half-power bandwidth of the low-pass filter. f_B can be calculated from f_A and A or from B_N using

$$f_B = (A^{-2} - 1)^{-1/2N} f_A \tag{4}$$

or

$$f_B = B_N \,\frac{\sin\left(\frac{\pi}{2N}\right)}{\left(\frac{\pi}{2N}\right)} \tag{5}$$

2. Tchebychev low-pass filters. The poles of a Tchebychev low-pass filter with N poles will be located at

$$P_k = -\,\omega_T \sinh\left[N^{-1}\sinh^{-1}(\epsilon^{-1})\right] \sin\left[(2k-1)\frac{\pi}{2N}\right] - i\omega_T \cosh\left[N^{-1}\sinh^{-1}(\epsilon^{-1})\right] \cos\left[(2k-1)\frac{\pi}{2N}\right], \qquad 1 \leqslant k \leqslant N \tag{6}$$

where $f_T = (2\pi)^{-1}\omega_T$ is the cut-off frequency of the Tchebychev filter and

$$\epsilon = (\rho^{-2} - 1)^{1/2} \tag{7}$$

where $20 \log_{10}(\rho)$ is the Tchebychev filter ripple factor. When f_A is specified, f_T can be computed using

$$f_T = \frac{1}{\cosh\left\{N^{-1}\cosh^{-1}\left[\epsilon^{-1}(A^{-2}-1)^{1/2}\right]\right\}}\, f_A, \qquad N = \text{odd}$$

$$= \frac{1}{\cosh\left\{N^{-1}\cosh^{-1}\left[\epsilon^{-1}(A^{-2}(1+\epsilon^2)-1)^{1/2}\right]\right\}}\, f_A, \qquad N = \text{even} \tag{8}$$

When B_N is specified, the poles and residues, P'_k and Q'_k, $1 \leqslant k \leqslant N$, of a Tchebychev filter with unit cutoff frequency ($f_T = 1$) are calculated. Then the poles and residues of the desired filter are

$$P_k = \frac{B_N}{B'_N}\, P'_k, \qquad 1 \leqslant k \leqslant N \tag{9}$$

and

$$Q_k = \frac{B_N}{B'_N}\, Q'_k, \qquad 1 \leqslant k \leqslant N \tag{10}$$

where

$$B'_N = -\frac{1}{4}\sum_{k=1}^{N}\left(\frac{Q'_k}{P'_k}\right)^2 - \sum_{k=1}^{N}\sum_{\ell=1}^{k-1}\left(\frac{Q'_k\,Q'_\ell}{P'_k + P'_\ell}\right) \tag{11}$$

is the noise bandwidth of the Tchebychev filter with unit cutoff frequency.

3. Bessel low-pass filter. An N-pole Bessel low-pass filter with unit response at zero frequency has transfer function

$$H_{LP}(s) = \frac{a_{NO}}{p_N(cs)} \tag{12}$$

where

$$p_N(s) = \sum_{k=0}^{N} a_{Nk}\, s^k \tag{13}$$

is the Nth order Bessel polynomial and c is a parameter which is adjusted to give either the required bandwidth f_A at attenuation A or the required noise bandwidth B_N. The Nth order Bessel polynomial is defined by the recursion formula

$$p_N(s) = (2N-1)p_{N-1}(s) + s^2 p_{N-2}(s) \tag{14}$$

where $p_0(s) = 1$ and $p_1(s) = s + 1$. The corresponding recursion relationships for the Bessel polynomial coefficients are

$$a_{N0} = (2N-1)\,a_{N-1,0} \tag{15}$$

$$a_{N1} = (2N-1)\,a_{N-1,1} \tag{16}$$

$$a_{Nk} = (2N-1)\,a_{N-1,k} + a_{N-2,k-2}, \qquad 2 \leqslant k \leqslant N-1 \tag{17}$$

and

$$a_{NN} = A_{N-2,N-2} \tag{18}$$

Used recursively, Eq. (12) and Eqs. (15) through (18) provide a simple method of calculating (for $c = 1$) the coefficients in the transfer function of a Bessel filter of any number of poles.

When f_A and A are specified, one first adjusts c to obtain the desired bandwidth and then employs a polynomial root funding subroutine to calculate the filter poles (the roots of $p_N(cs)/a_{N0}$). First, one finds the frequency $f_A' = (2\pi)^{-1}\omega_A'$ at which $|a_{N0}/p_N(s)|$ is equal to A. This is equivalent to finding the zero of

$$I(\omega) = 1 - A^2 \left\{ \left[1 + \sum_{\substack{k=2 \\ k=\text{even}}}^{N} (-1)^{k/2} \frac{a_{Nk}}{a_{N0}} \omega^k \right]^2 + \left[\sum_{\substack{k=1 \\ k=\text{odd}}}^{N} (-1)^{(k-1)/2} \frac{a_{Nk}}{a_{N0}} \omega^k \right]^2 \right\} \tag{19}$$

This can be solved using a series of linear extrapolations. To use this method an expression for $I'(\omega) = dI(\omega)/d\omega$ is needed. Differentiating Eq. (19),

$$I'(\omega) = -2A^2 \left\{ \left[1 + \sum_{\substack{k=2 \\ k=\text{even}}}^{N} (-1)^{k/2} \frac{a_{Nk}}{a_{N0}} \omega^k \right] \times \left[\sum_{\substack{k=2 \\ k=\text{even}}}^{N} (-1)^{k/2} \frac{a_{Nk}}{a_{N0}} k\omega^{k-1} \right] + \left[\sum_{\substack{k=1 \\ k=\text{odd}}}^{N} (-1)^{(k-1)/2} \frac{a_{Nk}}{a_{N0}} \omega^k \right] \times \left[\sum_{\substack{k=1 \\ k=\text{odd}}}^{N} (-1)^{(k-1)/2} \frac{a_{Nk}}{a_{N0}} k\omega^{k-1} \right] \right\} \tag{20}$$

Having determined ω_A' such that $I(\omega_A') = 0$,

$$c = \frac{\omega_A'}{\omega_A} \tag{21}$$

When the noise bandwidth B_N is specified, a polynomial root-funding subroutine can be used to calculate the poles P_K', $1 \leqslant k \leqslant N$, of $a_{N0}/p_N(s)$. Then, the residues Q_k' of $a_{N0}/p_N(s)$ at these poles can be calculated using P_K' instead of P_K in Eq. (2). Finally, Eqs. (9), (10) and (11) above can be used to calculate the poles and residues of a Bessel filter with the desired noise bandwidth.

4. **High-pass filters**. A high-pass filter can be synthesized from a low-pass filter using the transformation

$$H_{HP}(s) = H_{LP}\left(\frac{\omega_A^2}{s}\right) \tag{22}$$

If the low-pass filter has unit response at zero frequency, amplitude response A at frequency f_A, and zero response at infinite frequency, the high-pass filter will have zero response at zero frequency, amplitude response A at frequency f_A, and unit response at infinite frequency. If P_K and Q_K are the poles and residues of the low-pass filter, the response at infinity of the high-pass filter is

$$Q_0' = -\sum_{k=1}^{N} \frac{Q_k}{P_k} \tag{23}$$

The location of the kth pole of the high-pass filter will be

$$P'_k = \frac{\omega_A^2}{P_k} \tag{24}$$

and the residue at that pole will be

$$Q'_k = -\frac{\omega_A^2}{P_k^2} Q_k \tag{25}$$

5. **Band-pass filters**. A band-pass filter can be synthesized from a low-pass filter using the transformation

$$H_{BP} = H_{LP}\left(s + \frac{\omega_R^2}{s}\right) \tag{26}$$

where $f_R = (2\pi)^{-1}\omega_R$ is the band-pass filter resonant frequency. The resonant frequency will be approximately equal to the filter center frequency. If the low-pass filter has unit response at zero frequency and zero response at infinite frequency, the band-pass filter will have zero response at zero frequency, unit response at f_R, and zero response at infinite frequency. Furthermore, if the low-pass filter has attenuation A at frequency f_A, the band-pass filter will have bandwidth f_A between points at which the band-pass filter attenuation is A.

If P_K and Q_K, $1 \leqslant k \leqslant N$, are the poles and residues of the low-pass filter, the response at infinity of the band-pass filter will be zero and the band-pass filter poles will be

$$P'_k = \frac{1}{2} P_k + i\omega_R \left[1 - \left(\frac{P_k}{2\omega_R}\right)^2\right]^{1/2}, \qquad 1 \leqslant k \leqslant N \tag{27}$$

and

$$P''_k = \frac{1}{2} P_k - i\omega_R \left[1 - \left(\frac{P_k}{2\omega_R}\right)^2\right]^{1/2}, \qquad 1 \leqslant k \leqslant N \tag{28}$$

Note that the band-pass filter has $2N$ poles. The residues at these poles will be

$$Q'_k = Q_k \left(1 - \frac{P''_k}{P'_k}\right)^{-1} \tag{29}$$

and

$$Q''_k = Q_k \left(1 - \frac{P'_k}{P''_k}\right)^{-1} \tag{30}$$

6. **Band-reject filters**. A band-reject filter can be synthesized from a low-pass filter using the transformation

$$H_{BR}(s) = H_{LP}\left[\omega_A^2 \left(s + \frac{\omega_R^2}{s}\right)^{-1}\right] \tag{31}$$

If the low-pass filter has unit response at zero frequency and zero response at infinite frequency, the band-reject filter will have unit response at zero frequency, zero response at f_R, and unit response at infinite frequency. Furthermore, if the low-pass filter has attenuation A at frequency f_A, the band-reject filter will have bandwidth f_A between points at which the band-reject filter attenuation is A.

If P_K and Q_K, $1 \leqslant k \leqslant N$, are the poles and residues of the low-pass filter, the response at infinity of the band-reject filter will be

$$Q'_0 = -\sum_{k=1}^{N} \frac{Q_k}{P_k} \tag{32}$$

and the band-reject filter poles will be

$$P'_k = \frac{\omega_A^2}{2P_k} + i\omega_R \left[1 - \left(\frac{\omega_A^2}{2P_k\omega_R}\right)^2\right]^{1/2}, \qquad 1 \leqslant k \leqslant N \tag{33}$$

and

$$P''_k = \frac{\omega_A^2}{2P_k} - i\omega_R \left[1 - \left(\frac{\omega_A^2}{2P_k\omega_R}\right)^2\right]^{1/2}, \qquad 1 \leqslant k \leqslant N \tag{34}$$

Note that the band-reject filter has $2N$ poles. The residues at these poles will be

$$Q'_k = -\frac{\omega_A^2}{P_k^2} Q_k \left(1 - \frac{P''_k}{P'_k}\right)^{-1} \tag{35}$$

and

$$Q''_k = -\frac{\omega_A^2}{P_k^2} Q_k \left(1 - \frac{P'_k}{P''_k}\right)^{-1} \tag{36}$$

7. **Cascaded elements.** Having developed equations for calculating the response at infinity, the poles, and the residues at these poles for the individual elements of the RFI filter, the next step is to calculate the response at infinity, the poles, and the residues at these poles for the combination of a series of such elements. Consider two filters, one with N_1 poles and another with N_2 poles, which have no common pole. Let Q_{01} be the response at infinity and P_{1K} and Q_{1K} be the poles and residues of the first filter and Q_{02} be the response at infinity and P_{2K} and Q_{2K} be the poles and residues for the second filter. The cascade of the two filters has $N_1 + N_2$ poles, P_{1K}, $1 \leqslant k \leqslant N_1$, and P_{2K}, $1 \leqslant k \leqslant N_2$. The response at infinity of the cascaded filter will be

$$Q_0' = Q_{01} Q_{02} \tag{37}$$

The cascaded filter residue at P_{1K} will be

$$Q_{1k}' = Q_{1k} \left[Q_{02} + \sum_{\ell=1}^{N_2} \frac{Q_{2\ell}}{P_{1k} - P_{2\ell}} \right] \tag{38}$$

and that at P_{2K} will be

$$Q_{2k}' = Q_{2k} \left[Q_{01} + \sum_{\ell=1}^{N_1} \frac{Q_{1\ell}}{P_{2k} - P_{1\ell}} \right] \tag{39}$$

Starting with two RFI filter elements, the above process can be repeated until the response of the RFI filter at infinity, the RFI filter poles, and the residues at these poles have been calculated.

B. RF Receiving System

Figure 2 shows the analytical model for the demodulation of the telemetry data signal from the RF carrier. Note that noise is introduced at two different locations in Fig. 2. $n_0(t)$ represents the noise, such as atmospheric and antenna noise, introduced prior to the RFI filter. $n_2(t)$ represents the noise, such as the FET or maser LNA noise, introduced after the RFI filter. Two noise sources are required because the RFI filter will reduce the effective telemetry detector noise bandwidth for noise introduced prior to the RFI filter, but not for noise introduced after the RFI filter. The analysis assumes that the one-sided power spectral densities of $n_0(t)$ and $n_2(t)$ are $(1 - \eta)N_0$ and ηN_0 where N_0 is the one-sided power spectral density of the receiving system noise produced by all sources and η, $0 \leqslant \eta \leqslant 1$, is the fraction of this noise introduced after the RFI filter.

Referring again to Fig. 2, the signal and noise components of the RFI filter input will be

$$s_0(t) = (2P)^{1/2} \sin \left[\omega_C t + \phi_C + \beta x(t)\right] \tag{40}$$

and

$$n_0(t) = n_{0I}(t)\, 2^{1/2} \sin(\omega_C t + \phi_C - \phi)$$

$$+ n_{0Q}(t)\, 2^{1/2} \cos(\omega_C t + \phi_C - \phi) \tag{41}$$

P is the received signal power, $f_C = (2\pi)^{-1}\omega_C$ is the carrier frequency, and ϕ_C is the carrier phase. β is the telemetry modulation level and $x(t)$ is the binary-valued (± 1) telemetry modulation. This analysis assumes $\overline{x(t)}$ is zero. $n_{0I}(t)$ and $n_{0Q}(t)$ are statistically independent gaussian processes with constant one-sided power spectral density $(1 - \eta)N_0$ over the frequency band occupied by the telemetry data signal. Note that for analytical convenience the quadrature noise components of $n_0(t)$ are referenced to a carrier with instantaneous phase $\omega_C t + \phi_C - \phi$, where ϕ is the static phase error in the RF receiver phase-locked loop.

For the input signal and noise specified in Eqs. (40) and (41), the signal and noise output of an RFI filter with impulse response $h(t)$, amplitude response $A(\omega)$, and phase response $\phi(\omega)$ will be

$$s_1(t) = (2P)^{1/2} \cos(\beta) A(\omega_C) \sin\left[\omega_C t + \phi_C + \phi(\omega_C)\right]$$

$$+ (2P)^{1/2} \sin(\beta) \cos\left[\omega_C t + \phi_C + \phi(\omega_C)\right]$$

$$\times \int_0^\infty h_C(\tau)\, x(t - \tau)\, d\tau$$

$$+ (2P)^{1/2} \sin(\beta) \sin\left[\omega_C t + \phi_C + \phi(\omega_C)\right]$$

$$\times \int_0^\infty h_S(\tau)\, x(t - \tau)\, d\tau \tag{42}$$

and

$$n_1(t) = 2^{1/2} \sin [\omega_C t + \phi_C + \phi(\omega_C) - \phi]$$

$$\times \int_0^\infty h_C(\tau)\, n_{0I}(t-\tau)\, d\tau$$

$$- 2^{1/2} \cos [\omega_C t + \phi_C + \phi(\omega_C) - \phi]$$

$$\times \int_0^\infty h_S(\tau)\, n_{0I}(t-\tau)\, d\tau$$

$$+ 2^{1/2} \cos [\omega_C t + \phi_C + \phi(\omega_C) - \phi]$$

$$\times \int_0^\infty h_C(\tau)\, n_{0Q}(t-\tau)\, d\tau$$

$$+ 2^{1/2} \sin [\omega_C t + \phi_C + \phi(\omega_C) - \phi]$$

$$\times \int_0^\infty h_S(\tau)\, n_{0Q}(t-\tau)\, d\tau \qquad (43)$$

where

$$h_C(\tau) = h(\tau) \cos [\omega_C \tau + \phi(\omega_C)] \qquad (44)$$

and

$$h_S(\tau) = h(\tau) \sin [\omega_C \tau + \phi(\omega_C)] \qquad (45)$$

The noise $n_2(t)$ introduced after the RFI filter will have the form

$$n_2(t) = n_{2I}(t)\, 2^{1/2} \sin [\omega_C t + \phi_C + \phi(\omega_C) - \phi]$$

$$+ n_{2Q}(t)\, 2^{1/2} \cos [\omega_C t + \phi_C + \phi(\omega_C) - \phi] \qquad (46)$$

where, as in Eq. (41), the carrier instantaneous phase to which the quadrature noise components are referenced is chosen for analytical convenience.

Assuming the RF receiver phase-locked loop tracks the carrier component of the RFI filter output with phase error ϕ, the reference signal for carrier demodulation shown in Fig. 2 will be

$$s_3(t) = 2^{1/2} \cos [\omega_C t + \phi_C + \phi(\omega_C) - \phi] \qquad (47)$$

Thus, neglecting terms with zero frequency and terms in $2\omega_C t$, the signal and noise outputs of the carrier demodulator will be

$$s_4(t) = P^{1/2} \sin (\beta) \int_0^\infty [h_C(\tau) \cos (\phi)$$

$$+ h_S(\tau) \sin (\phi)]\, x(t-\tau)\, d\tau \qquad (48)$$

and

$$n_4(t) = n_{2Q}(t) + \int_0^\infty h_C(\tau)\, n_{0Q}(t-\tau)\, d\tau$$

$$- \int_0^\infty h_S(\tau)\, n_{0I}(t-\tau)\, d\tau \qquad (49)$$

Using the above results, Fig. 3 shows a simplified analytical model for the DSN RF receiving system and telemetry detector. The transfer functions $H'(s)$, $H_C(s)$, and $H_S(s)$ shown in Fig. 3 are those corresponding to the impulse responses $h'(\tau)$, $h_C(\tau)$, and $h_S(\tau)$ where

$$h'(\tau) = h_C(\tau) \cos (\phi) + h_S(\tau) \sin (\phi) \qquad (50)$$

For an RFI filter having a transfer function of the form shown in Eq. (1),

$$H_C(s) = Q_{C0} + \sum_{k=1}^{N} \frac{Q_{Ck}}{s - P_{Ck}} \qquad (51)$$

and

$$H_S(s) = Q_{S0} + \sum_{k=1}^{N} \frac{Q_{Sk}}{s - P_{Sk}} \qquad (52)$$

where

$$P_{Ck} = P_{Sk} = P_k - i\omega_C, \qquad Im(P_k) > 0$$

$$= P_k + i\omega_C, \qquad Im(P_k) < 0 \qquad (53)$$

where $Im(z)$ denotes the imaginary part of a complex variable z,

$$Q_{Ck} = \frac{1}{2} Q_k \exp\left[-i\phi(\omega_C)\right], \qquad Im(P_k) > 0$$

$$= \frac{1}{2} Q_k \exp\left[+i\phi(\omega_C)\right], \qquad Im(P_k) < 0 \qquad (54)$$

$$Q_{Sk} = -\frac{1}{2i} Q_k \exp\left[-i\phi(\omega_C)\right], \qquad Im(P_k) > 0$$

$$= \frac{1}{2i} Q_k \exp\left[i\phi(\omega_C)\right], \qquad Im(P_k) < 0 \qquad (55)$$

$$Q_{C0} = Q_0 \cos\left[\phi(\omega_C)\right] \qquad (56)$$

and

$$Q_{S0} = Q_0 \sin\left[\phi(\omega_C)\right] \qquad (57)$$

Thus,

$$H'(s) = Q_0' + \sum_{k=1}^{N} \frac{Q_k'}{S - P_k'} \qquad (58)$$

where

$$P_k' = P_{Ck} = P_{Sk}, \qquad 1 \leqslant k \leqslant N \qquad (59)$$

and

$$Q_k' = Q_{Ck} \cos(\phi) + Q_{Sk} \sin(\phi), \qquad 0 \leqslant k \leqslant N \qquad (60)$$

Equations (53) through (57) and Eqs. (59) and (60) provide a means of calculating the response at infinity and the poles and residues of the transfer functions $H'(s)$, $H_C(s)$, and $H_S(s)$ shown in Fig. 3, given the response at infinity and the poles and residues of the RFI filter.

The effect of the *noise* filtering shown in Fig. 3 on telemetry system performance will be to increase telemetry performance by $1/\delta_N$ where δ_N is the factor by which the variance of the noise at the integrate-and-dump circuit output is reduced by the filtering. The next section of this report provides a means of calculating the ratio of the bandwidth of an integrate-and-dump circuit preceded by a filter with only simple poles to that of the integrate-and-dump circuit alone. Thus, if δ_{NC} is this ratio for the filter with transfer function $H_C(s)$ and δ_{NS} is this ratio for the filter with transfer function $H_S(s)$,

$$\delta_N = \eta + (1 - \eta)(\delta_{NC} + \delta_{NS}) \qquad (61)$$

Thus, as expected, if all of the receiving system noise is introduced after the RFI filter, $\eta = 1$ and $\delta_N = 1$. When the noise is introduced after the RFI filter, the RFI filter does not reduce the effect of the noise on telemetry performance.

C. Reduction in the Telemetry Detector Noise Bandwidth Produced by a Filter Preceding the Integrate-and-Dump Circuit

As discussed in the preceding section, a method is needed for calculating the reduction in telemetry detector noise bandwidth caused by a filter with only simple poles preceding the integrate-and-dump circuit. Since the filter has only simple poles, its impulse response will have the form

$$h_1(\tau) = Q_0\delta(\tau) + \sum_{k=1}^{N} Q_k \exp(P_k\tau) \qquad (62)$$

where Q_0 is the filter response at infinite frequency, P_K and Q_k, $1 \leqslant k \leqslant N$, are the filter poles and residues, and $\delta(\tau)$ is the Dirac delta function. The impulse response of the integrate-and-dump circuit will be

$$h_2(\tau) = 1, \qquad 0 \leqslant \tau \leqslant T,$$

$$= 0, \qquad \tau > T \qquad (63)$$

where T is the integration time. The impulse response of the cascaded filters will be the convolution of $h_1(t)$ and $h_2(t)$ or

$$h(\tau) = Q_0 + \sum_{k=1}^{N} \frac{Q_k}{P_k}\left[\exp(P_k\tau) - 1\right], \qquad 0 \leqslant \tau \leqslant T$$

$$= \sum_{k=1}^{N} \frac{Q_k}{P_k}\left[1 - \exp(-P_kT)\right]\exp(P_k\tau), \qquad T \leqslant \tau \leqslant \infty$$

$$(64)$$

The one-sided noise bandwidth of a filter with impulse response $h(\tau)$ will be

$$B_N = \frac{1}{2} \int_0^\infty h^2(\tau)\, d\tau \tag{65}$$

The noise bandwidth of the integrate-and-dump circuit above is $T/2$. Thus, using Eq. (64) in Eq. (65), the reduction in noise bandwidth caused by the filtering in front of the integrate-and-dump circuit will be

$$\frac{B_N}{T/2} = \left(Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right)^2 + 2 \sum_{k=1}^{N} \frac{Q_k}{P_k} \frac{\exp(P_k T) - 1}{P_k T} \left(Q_0 - \sum_{\ell=1}^{N} \frac{Q_\ell}{P_k + P_\ell}\right) \tag{66}$$

Having computed the response at infinity and the poles and residues for the transfer functions $H_C(s)$ and $H_S(s)$, using the equations derived in the previous section, Eq. (66) with appropriate substitutions, can be used to calculate δ_{NC} and δ_{NS}.

D. Probability Density of the Integrate-and-Dump Circuit Output Signal Level

As a result of the filter with transfer function $H'(s)$ shown in Fig. 3, the signal integrated by the integrate-and-dump circuit will depend not only on the data bit being detected, but also on preceding bits and, in this analysis, on one succeeding bit. The dependence on preceding bits occurs because the impulse response of $H'(s)$ will, in general, be of infinite duration; that is, the current output of $H'(s)$ depends on all previous inputs. The dependence on the succeeding bit is a consequence of the displacement in bit synchronizer timing caused by $H'(s)$. This analysis assumes the displacement in bit synchronizer timing does not exceed one bit.

Each different signal waveform that can occur during a bit integration can result in a different integrate-and-dump circuit output signal level. This analysis assumes that the impulse response of $H'(s)$ is significant for only M bit times. If the impulse response extends over M preceding bits, and one succeeding bit is significant because of bit synchronizer timing displacement, there are 2^{M+2} possible integrate-and-dump circuit output levels. This analysis assumes that all data sequences occur with equal probability. Hence, all of the 2^{M+2} possible integrate-and-dump circuit output signal levels occur with equal probability $1/2^{M+2}$.

The next step is to calculate each of the 2^{M+2} possible output signal levels. Refer to the simplified block diagram in Fig. 4 and the sample $x(t)$ waveform in Fig. 5. Note that the nth data bit in $x(t)$ is received during $[(\lambda + n - 1)T, (\lambda + n)T]$. Assume that because of the distortion of $x(t)$ the bit synchronizer causes the integrate-and-dump circuit to integrate over $[(\hat{\lambda} + n - 1)T, (\hat{\lambda} + n)T]$, during detection of the nth data bit, where $\hat{\lambda} = \lambda + \epsilon$. Then

$$z_n = \int_{(\lambda+n-1+\epsilon)T}^{(\lambda+n+\epsilon)T} y(t)\, dt \tag{67}$$

or

$$z_n = \int_{(\lambda+n-1+\epsilon)T}^{(\lambda+n)T} y(t)\, dt + \int_{(\lambda+n)T}^{(\lambda+n+\epsilon)T} y(t)\, dt, \qquad 0 \leqslant \epsilon \leqslant 1 \tag{68}$$

and

$$z_n = \int_{(\lambda+n-1+\epsilon)T}^{(\lambda+n-1)T} y(t)\, dt + \int_{(\lambda+n-1)T}^{(\lambda+n+\epsilon)T} y(t)\, dt, \qquad -1 \leqslant \epsilon \leqslant 0 \tag{69}$$

If $h'(\tau)$, the impuse response corresponding to the transfer function $H'(s)$, has the form

$$h'(\tau) = Q_0 \delta(\tau) + \sum_{k=1}^{N} Q_k \exp(P_k \tau) \tag{70}$$

$$\begin{aligned} y(t) = {} & x_n \sum_{k=1}^{N} \frac{Q_k}{P_k} \left[\exp\{P_k [t - (\lambda + n - 1)T]\} - 1\right] \\ & + \sum_{m=-\infty}^{n-1} x_m \sum_{k=1}^{N} \frac{Q_k}{P_k} \\ & \times [\exp\{P_k [t - (\lambda + m - 1)T]\} \\ & - \exp\{P_k [t - (\lambda + m)T]\}], \\ & (\lambda + n - 1)T \leqslant t \leqslant (\lambda + n)T \end{aligned} \tag{71}$$

Using Eq. (71) in Eqs. (68) and (69),

$$z_n = \sum_{m=-1}^{\infty} c_m x_{n-m}, \qquad x_{n-m} = \pm 1 \tag{72}$$

where

$$c_0 = (1-\epsilon)\left[Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right]$$

$$+ \sum_{k=1}^{N} \frac{Q_k}{P_k^2 T} \{\exp[(1+\epsilon)P_k T]$$

$$- 2\exp(\epsilon P_k T) + 1\}, \quad 0 \leqslant \epsilon \leqslant 1 \qquad (73)$$

$$c_0 = (1+\epsilon)\left[Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right]$$

$$+ \sum_{k=1}^{N} \frac{Q_k}{P_k^2 T} \{\exp[(1+\epsilon)P_k T] - 1\}, \quad -1 \leqslant \epsilon \leqslant 0 \qquad (74)$$

$$c_{-1} = \epsilon\left[Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right]$$

$$+ \sum_{k=1}^{N} \frac{Q_k}{P_k^2 T} [\exp(\epsilon P_k T) - 1], \quad 0 \leqslant \epsilon \leqslant 1 \qquad (75)$$

$$c_{-1} = 0, \quad -1 \leqslant \epsilon \leqslant 0 \qquad (76)$$

$$c_1 = \sum_{k=1}^{N} \frac{Q_k}{P_k^2 T} \{\exp[(2+\epsilon)P_k T]$$

$$-2\exp[(1+\epsilon)P_k T] + \exp(\epsilon P_k T)\}, \quad 0 \leqslant \epsilon \leqslant 1 \qquad (77)$$

$$c_1 = -\epsilon\left[Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right]$$

$$+ \sum_{k=1}^{N} \frac{Q_k}{P_k^2 T} \{\exp[(2+\epsilon)P_k T]$$

$$- 2\exp[(1+\epsilon)P_k T] + 1\}, \quad -1 \leqslant \epsilon \leqslant 0 \qquad (78)$$

and

$$c_m = \sum_{k=1}^{N} \frac{Q_k}{P_k^2 T} \{\exp[(m+1+\epsilon)P_k T] - 2\exp[(m+\epsilon)P_k T]$$

$$+ \exp[(m-1+\epsilon)P_k T]\}, \quad -1 \leqslant \epsilon \leqslant 1, m \geqslant 2 \qquad (79)$$

Thus, approximating the infinite summation in Eq. (72) with one whose upper limit is M, the probability density function of the integrate-and-dump circuit output signal level, normalized to $P^{1/2}\sin(\beta)$, is

$$p(z_n) = \frac{1}{2^{M+2}} \sum_{x_{n-M}=\pm 1} \cdots \sum_{x_{n+1}=\pm 1}$$

$$\times\, \delta\left(z_n - \sum_{m=-1}^{M} x_{n-m} c_m\right) \qquad (80)$$

E. Telemetry Performance Degradation

Having determined the probability density function of the integrate-and-dump circuit output signal level and the reduction in the noise bandwidth of the telemetry detector, it remains to calculate the telemetry performance degradation. In a system with no filter or RF receiver phase error ($\phi = 0$), the channel bit error probability will be

$$P_B = \frac{1}{2}\left\{1 - Erf\left[\left(\frac{E_B}{N_0}\right)^{1/2}\right]\right\} \qquad (81)$$

where E_B is the received signal energy per bit and N_0 is the receiving system noise spectral density. In such a system, z_n and δ_N will be one. When z_n and δ_N are not one, the effect on the bit error probability is equivalent to that of reducing $(E_B/N_0)^{1/2}$ by $z_N/\delta_N^{1/2}$. Therefore, in the presence of signal distortion and RF receiver phase error, the bit error probability will be

$$P_B = \frac{1}{2^{M+1}} \sum_{\mathbf{x}} \frac{1}{2}\left\{1 - Erf\left[\left(\frac{E_B}{N_0} \cdot \frac{1}{\delta_N}\right)^{1/2} (\mathbf{x} \cdot \mathbf{c})\right]\right\} \qquad (82)$$

x is the vector $(x_{n-M}, \ldots, x_{n-1}, 1, x_{n+1})$, **c** is the vector $(c_M, \ldots, c_{-1})$, **x** · **c** is the dot product of the two vectors, and

$$\sum_{\mathbf{x}}$$

denotes the summation over all the 2^{M+1} possible values of x. Note that to simplify the above analysis (without any loss of generality) the nth transmitted bit was assumed to be a binary zero ($x_n = 1$). If P_B is the allowable telemetry channel bit error probability and $(E_B/N_0)_B$ is the E_B/N_0 that yields this bit error probability in the absence of filtering and RF receiver phase error, the telemetry channel performance degradation δ_T will be the solution of

$$P_B = \frac{1}{2^{M+1}} \sum_{\mathbf{x}} \frac{1}{2}\left[1 - Erf\left\{\left[\left(\frac{E_B}{N_0}\right)_B \cdot \frac{\delta_T}{\delta_N}\right]^{1/2} (\mathbf{x}\cdot\mathbf{c})\right\}\right] \tag{83}$$

This equation can be solved for δ_T using a series of linear extrapolations starting with some initial estimate for δ_T (i.e., $\delta_T = 1$). Since P_B will be almost an exponential function of δ_T, convergence of this iterative process will be accelerated by working with the natural logarithm $\ln(x)$ of both sides of Eq. (83). Thus, the slope needed for the linear extrapolation will be

$$\frac{d\ln(P_B)}{d\delta_T} = \frac{1}{P_B}\frac{dP_B}{d\delta_T} \tag{84}$$

where

$$\frac{dp_B}{d\delta_T} = -\frac{1}{2}\left[\frac{\left(\frac{E_B}{N_0}\right)_B}{\pi\delta_N\delta_T}\right]^{1/2}$$

$$\times \frac{1}{2^{M+1}} \sum_{\mathbf{x}} (\mathbf{x}\cdot\mathbf{c}) \exp\left[-\frac{\left(\frac{E_B}{N_0}\right)_B \delta_T}{\delta_N} (\mathbf{x}\cdot\mathbf{c})^2\right] \tag{85}$$

F. Bit Synchronizer Timing Displacement

Examining Eqs. (82) and (83) and Eqs. (73) through (79), one notes that the telemetry channel bit error probability and the performance degradation δ_T will be a function of ϵ, the displacement in bit synchronizer timing expressed as a fraction of the bit time. While analytical prediction of the bit synchronizer timing displacement appears feasible, such an analysis would be quite difficult. In this report two alternate approaches are used.

1. Narrowband approximation. The first and simplest approach is to calculate an estimate of ϵ from the phase response of the filter with transfer function $H'(s)$ shown in Figs. 3 and 4. For narrowband telemetry modulation, the displacement in bit synchronizer timing will be equal to the time delay through $H'(s)$ at very low frequencies. Thus, if

$$H'(i\omega) = A'(\omega) \exp\left[i\phi'(\omega)\right] \tag{86}$$

the narrowband approximation to the displacement in bit synchronizer timing will be

$$\epsilon_I = \frac{1}{T}\frac{d\phi'(\omega)}{d\omega}\bigg|_{\omega=0} \tag{87}$$

If the response at infinity Q_0 and the poles and residues, P_k and Q_k $(1 \leqslant k \leqslant N)$ of $H'(s)$ are known, and

$$a_k = [R(P_k)]^2 + [I(P_k)]^2 \tag{88}$$

$$b_k = -2I(P_k) \tag{89}$$

$$c_k = -R(Q_k)R(P_k) - I(Q_k)I(P_k) \tag{90}$$

$$d_k = I(Q_k) \tag{91}$$

$$e_k = R(Q_k)I(P_k) - I(Q_k)R(P_k) \tag{92}$$

$$f_k = -R(Q_k) \tag{93}$$

where $R(z)$ and $I(z)$ denote the real and imaginary parts of a complex variable z,

$$\epsilon_I = \frac{1}{T}\frac{X(0)Y'(0) - Y(0)X'(0)}{[X(0)]^2 + [Y(0)]^2} \tag{94}$$

where

$$X(0) = Q_0 + \sum_{k=1}^{N} \frac{c_k}{a_k} \tag{95}$$

$$Y(0) = \sum_{k=1}^{N} \frac{e_k}{a_k} \tag{96}$$

$$X'(0) = \sum_{k=1}^{N} \frac{(a_k d_k - b_k c_k)}{a_k^2} \tag{97}$$

and

$$Y'(0) = \sum_{k=1}^{N} \frac{(a_k f_k - b_k e_k)}{a_k^2} \tag{98}$$

2. Bit synchronizer timing displacement for minimum telemetry degradation. The second and somewhat more complex approach is to find the bit synchronizer timing displacement which minimizes the telemetry performance degradation. If the bit synchronizer does not automatically adjust the timing to minimize the telemetry degradation, the timing could be manually adjusted to achieve the minimum performance degradation. Of course, requiring manual adjustment of bit synchronizer timing might create some operational difficulties.

A second iterative procedure, also using a series of linear extrapolations, can be employed to find the value of ϵ (ϵ_o) that minimizes δ_T. Finding the value of ϵ that minimizes δ_T is equivalent to finding the zero of $d\delta_T/d\epsilon$. To find a zero crossing of a function, the iterative process requires an initial estimate of the solution and a means of calculating the function and its derivative. The initial estimate used herein (ϵ_I) is that obtained using the narrowband approximation discussed above. $d\delta_T/d\epsilon$ and $d^2\delta_T/d\epsilon^2$ can be calculated using

$$\frac{d\delta_T}{d\epsilon} = -2\delta_T \frac{\alpha}{\beta} \tag{99}$$

and

$$\frac{d^2\delta_T}{d\epsilon^2} = -2\left[\frac{\alpha}{\beta}\frac{d\delta_T}{d\epsilon} + \frac{\delta_T}{\beta}\frac{d\alpha}{d\epsilon} - \frac{\delta_T \alpha}{\beta^2}\frac{d\beta}{d\epsilon}\right] \tag{100}$$

where

$$\alpha = \sum_{\mathbf{x}} \left(\mathbf{x} \cdot \frac{d\mathbf{c}}{d\epsilon}\right) \exp\left[-\frac{\delta_T}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c})^2\right] \tag{101}$$

$$\beta = \sum_{\mathbf{x}} (\mathbf{x} \cdot \mathbf{c}) \exp\left[-\frac{\delta_T}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c})^2\right] \tag{102}$$

$$\begin{aligned} \frac{d\alpha}{d\epsilon} = \sum_{\mathbf{x}} &\left[\mathbf{x}\frac{d^2\mathbf{c}}{d\epsilon^2} - \frac{1}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c})^2 \left(\mathbf{x} \cdot \frac{d\mathbf{c}}{d\epsilon}\right)\frac{d\delta_T}{d\epsilon} \right. \\ &\left. - 2\frac{\delta_T}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c}) \left(\mathbf{x} \cdot \frac{d\mathbf{c}}{d\epsilon}\right)^2\right] \\ &\times \exp\left[-\frac{\delta_T}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c})^2\right] \end{aligned} \tag{103}$$

$$\begin{aligned} \frac{d\beta}{d\epsilon} = \sum_{\mathbf{x}} &\left[\mathbf{x} \cdot \frac{d\mathbf{c}}{d\epsilon} - (\mathbf{x} \cdot \mathbf{c})^3 \frac{1}{\delta_N}\left(\frac{E_B}{N_0}\right)_B \frac{d\delta}{d\epsilon} \right. \\ &\left. -2\frac{\delta_T}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c})^2 \left(\mathbf{x} \cdot \frac{d\mathbf{c}}{d\epsilon}\right)\right] \\ &\times \exp\left[-\frac{\delta_T}{\delta_N}\left(\frac{E_B}{N_0}\right)_B (\mathbf{x} \cdot \mathbf{c})^2\right] \end{aligned} \tag{104}$$

$$\begin{aligned} \frac{d\mathbf{c}_0}{d\epsilon} = -&\left[Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right] \\ &+ \sum_{k=1}^{N} \frac{Q_k}{P_k} \{\exp[(1+\epsilon)P_k T] \\ &-2\exp(\epsilon P_k T)\}, \quad 0 \leqslant \epsilon \leqslant 1 \end{aligned} \tag{105}$$

$$\begin{aligned} \frac{d^2\mathbf{c}_0}{d\epsilon^2} = &\sum_{k=1}^{N} Q_k T \{\exp[(1+\epsilon)P_k T] \\ &-2\exp(\epsilon P_k T)\}, \quad 0 \leqslant \epsilon \leqslant 1 \end{aligned} \tag{106}$$

$$\frac{dc_0}{d\epsilon} = Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k} + \sum_{k=1}^{N} \frac{Q_k}{P_k} \exp\,[(1+\epsilon)P_kT], \quad -1 \leqslant \epsilon \leqslant 0 \tag{107}$$

$$\frac{d^2c_0}{d\epsilon^2} = \sum_{k=1}^{N} Q_kT \exp\,[(1+\epsilon)P_kT], \quad -1 \leqslant \epsilon \leqslant 0 \tag{108}$$

$$\frac{dc_{-1}}{d\epsilon} = Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k} + \sum_{k=1}^{N} \frac{Q_k}{P_k} \exp\,(\epsilon P_kT), \quad 0 \leqslant \epsilon \leqslant 1 \tag{109}$$

$$\frac{d^2c_{-1}}{d\epsilon^2} = \sum_{k=1}^{N} Q_kT \exp\,(\epsilon P_kT), \quad 0 \leqslant \epsilon \leqslant 1 \tag{110}$$

$$\frac{dc_{-1}}{d\epsilon} = \frac{d^2c_{-1}}{d\epsilon^2} = 0, \quad -1 \leqslant \epsilon \leqslant 0 \tag{111}$$

$$\frac{dc_1}{d\epsilon} = \sum_{k=1}^{N} \frac{Q_k}{P_k} \{\exp\,[(2+\epsilon)P_kT] \; -2 \exp\,[(1+\epsilon)P_kT] + \exp\,(\epsilon P_kT)\}, \quad 0 \leqslant \epsilon \leqslant 1 \tag{112}$$

$$\frac{d^2c_1}{d\epsilon^2} = \sum_{k=1}^{N} Q_kT \{\exp\,[(2+\epsilon)P_kT] -2 \exp\,[(1+\epsilon)P_kT] + \exp\,(\epsilon P_kT)\}, \quad 0 \leqslant \epsilon \leqslant 1 \tag{113}$$

$$\frac{dc_1}{d\epsilon} = -\left[Q_0 - \sum_{k=1}^{N} \frac{Q_k}{P_k}\right] + \sum_{k=1}^{N} \frac{Q_k}{P_k} \{\exp\,[(2+\epsilon)P_kT] -2 \exp\,[(1+\epsilon)P_kT]\}, \quad -1 \leqslant \epsilon \leqslant 0 \tag{114}$$

$$\frac{d^2c_1}{d\epsilon^2} = \sum_{k=1}^{N} Q_kT \{\exp\,[(2+\epsilon)P_kT] -2 \exp\,[(1+\epsilon)P_kT]\}, \quad -1 \leqslant \epsilon \leqslant 0 \tag{115}$$

$$\frac{dc_m}{d\epsilon} = \sum_{k=1}^{N} \frac{Q_k}{P_k} \{\exp\,[(m+1+\epsilon)P_kT] -2 \exp\,[(m+\epsilon)P_kT] + \exp\,[(m-1+\epsilon)P_kT]\}, \quad -1 \leqslant \epsilon \leqslant 1, m \geqslant 2 \tag{116}$$

$$\frac{d^2c_m}{d\epsilon^2} = \sum_{k=1}^{N} Q_kT \{\exp\,[(m+1+\epsilon)P_kT] -2 \exp\,[(m+\epsilon)P_kT] + \exp\,[(m-1+\epsilon)P_kT] \,, \quad -1 \leqslant \epsilon \leqslant 1, m \geqslant 2 \tag{117}$$

III. Numerical Results

Data are presented in this section for the telemetry channel degradation for the S-band FET amplifier, S-band maser amplifier, and X-band maser amplifier RFI filters. For each of these RFI filters telemetry channel degradation is plotted as a function of the telemetry channel carrier frequency for telemetry channel data rates of 4 Mbps, 10 Mbps, and 20 Mbps. The 4 Mbps rate is the highest data rate currently planned. The data for the 10 Mbps and 20 Mbps data rates were included to show the impact of the RFI filters on future systems. Plotting the telemetry channel degradation as a function of carrier frequency shows the impact of the RFI filter on telemetry channels whose center frequency does not fall at the RFI filter center frequency.

Two sets of curves are shown for each RFI filter. One set assumes that all of the receiving system noise is introduced after the RFI filter ($\eta = 1$). The other set assumes that all of the receiving system noise is introduced before the RFI filter ($\eta = 0$). The telemetry degradation δ_{T1} that occurs when all the noise is introduced after the RFI filter will always be greater than the telemetry degradation δ_{T0} that occurs when all the noise is introduced after the RFI filter, because the RFI filter reduces the telemetry system noise bandwidth for noise introduced prior to the RFI filter. Given the two sets of data ($\eta = 1$ and $\eta = 0$) for each RFI filter, the telemetry degradation for any other value of η is

$$\delta_T = \frac{\delta_{T1}\,\delta_{T0}}{(1-\eta)\delta_{T1} + \eta\delta_{T0}} \tag{118}$$

A. S-band FET Amplifier RFI Filter

Figures 6 and 7 show the telemetry channel degradation as a function of carrier frequency for the S-band FET amplifier RFI filter. Figure 6 assumes all of the noise is introduced after the RFI filter ($\eta = 1$). Figure 7 assumes all of the noise is introduced prior to the RFI filter ($\eta = 0$). The results presented in Figs. 6 and 7 assume that the RF receiver phase-locked loop static phase error is zero and that the allowable telemetry channel bit error probability is 10^{-5}.

The data shown in Figs. 6 and 7 are for 7th and 9th order Tchebychev bandpass filters with 270 MHz bandwidth between points at which the filter attenuation is -70 dB and 0.01 dB ripple factor. The filter resonant frequency, which is approximately equal to the filter center frequency, is 2250 MHz.

The telemetry channel degradation shown in Figs. 6 and 7 increases significantly with data rate, even when the carrier frequency is at the center of the RFI filter passband. The degradation increases significantly as the edge of the passband is approached. While the filter amplitude response may cause significant degradation, it is suspected that much of the degradation shown in Figs. 6 and 7 results from the non-linear phase response of the Tchebychev filter. While the degradation shown in Figs. 6 and 7 for the 4 Mbps data rate, especially that for the 9th order filter, seems acceptable, it appears that the use of this filter will cause significant telemetry degradation at much higher bit rates. This degradation will be significant even for telemetry channels whose center frequency is at the RFI filter center frequency.

B. S-band Maser Amplifier RFI Filter

Figures 8 and 9 show the telemetry channel degradation as a function of carrier frequency for the S-band maser amplifier RFI filter. Figure 8 assumes that all of the noise is introduced after the RFI filter ($\eta = 1$). Figure 9 assumes all of the noise is introduced before the RFI filter ($\eta = 0$). The results presented in Figs. 8 and 9 assume the RF receiver phase-locked loop static phase error is zero and that the allowable telemetry channel bit error probability is 10^{-5}.

The data shown in Figs. 8 and 9 are for a 5th order Tchebychev bandpass filter with 60 MHz bandwidth between points at which the filter attenuation is -0.5 dB and 0.01 dB ripple factor. The filter resonant frequency is 2292 MHz.

As in the case of the S-band FET amplifier RFI filter, the telemetry channel degradation increases significantly with data rate, even when the carrier frequency is near the RFI filter center frequency. The telemetry degradation does not seem unreasonable for the current 4 Mbps maximum data rate, but the degradation at higher data rates may not be acceptable.

C. X-band Maser Amplifier RFI Filter

Figures 10 and 11 show the telemetry channel degradation as a function of carrier frequency for the X-band maser amplifier RFI filter. Figure 10 assumes that all of the noise is introduced after the RFI filter ($\eta = 1$). Figure 11 assumes that all of the noise is introduced before the RFI filter ($\eta = 0$). The results shown in Figs. 10 and 11 assume that the RF receiver phase-locked loop phase error is zero and that the allowable telemetry channel bit error probability is 10^{-5}.

The data shown in Figs. 10 and 11 are for a 5th order Tchebychev bandpass filter with 160 MHz bandwidth between the points at which the filter attenuation is -3 dB and either 0.01 or 0.05 dB ripple factor. The filter resonant frequency is 8450 MHz. The 0.05 dB ripple factor curves in Fig. 11 were omitted for the 4 and 10 Mbps data rates, because they were essentially the same as the 0.01 dB ripple factor curves for the corresponding cases.

As in the cases of the S-band FET amplifier RFI filter and the S-band maser amplifier RFI filter, the telemetry channel degradation increases significantly with data rate. However, the telemetry channel degradation caused by the X-band maser amplifier RFI filter seems to be appreciably less than that produced by the S-band RFI filters. For 4 Mbps the degradation for $\eta = 1$, the worst case, is less than 0.1 dB over essentially all of the 8400 to 8500 MHz frequency band. For 10 Mbps the degradation is less than 0.2 dB over almost all of this frequency band. Even at 20 Mbps, the degradation is only about 0.25 dB for a carrier located at the RFI filter center frequency. Finally, the Tchebychev filter ripple factor does not appear to strongly affect the telemetry channel degradation produced by the X-band maser amplifier RFI filter.

D. Effect of Bit Synchronizer Timing Displacement

As noted in Section II.F of this article, this analysis does not calculate the actual bit synchronizer timing displacement. The "narrowband" approximation is used to obtain an initial estimate of bit synchronizer timing displacement and then an iterative process is employed to find the bit synchronizer timing displacement that minimizes telemetry channel degradation. It is this minimum value of the telemetry channel degradation that is plotted in Figs. 6 through 11.

The assumption in plotting this minimum telemetry channel degradation is that bit synchronizer timing can always be manually adjusted to optimize performance. However, as manual adjustment of bit synchronizer timing may be very cumbersome, let us assess the impact of non-optimum bit synchronizer timing displacement by comparing both the initial estimate of bit synchronizer timing displacement, obtained using the narrowband approximation, with the optimum value of bit synchronizer timing displacement and the telemetry channel degradations for these two estimates of bit synchronizer timing displacement.

Figure 12 compares the initial estimate and optimum value of bit synchronizer timing displacement, and Fig. 13 compares the corresponding telemetry degradations as a function of carrier frequency for the S-band maser RFI filter. Clearly, for carrier frequencies near the edge of the RFI filter passband, there is a substantial difference between the initial estimate and the optimum value of bit synchronizer timing displacement and also a substantial difference between the degradations. The degradation for the initial estimate is significantly greater, for carrier frequencies near the edge of the RFI filter passband, than the minimum degradation. This suggests that the telemetry degradation could be somewhat greater than that shown in Figs. 6 through 11, if actual bit synchronizer timing displacement differs from the optimum value as much as the initial estimate does and manual adjustment is not feasible.

IV. Conclusion

This article has presented a method for determining the effects of telemetry channel filtering on the performance of a discrete carrier PCM/PM telemetry channel and has applied this technique to calculate numerical results for the S-band FET amplifier, S-band maser amplifier and X-band maser amplifier RFI filters being considered by the DSN. This method can be adapted to calculate telemetry channel degradation for a very wide variety of RFI filters. The method can be easily extended to filters other than the cascaded Butterworth, Tchebychev, and Bessel low-pass, high-pass and band-reject filters considered herein, as long as the RFI filter has only simple poles and its poles and the residues at these poles can be calculated.

The numerical results presented for the S-band FET amplifier, S-band maser amplifier, and X-band maser amplifier RFI filters show that the telemetry channel degradation should be negligible for bit rates less than the current 4 Mbps maximum capability of the BBA (baseband assembly) telemetry processor. However, at data rates significantly greater than 4 Mbps, the telemetry degradation produced by the RFI filters, especially the S-band RFI filters, may be unacceptable.

Reference

1. "Final RFI Report," Ford Aerospace and Communications Corporation, 12-7-81.

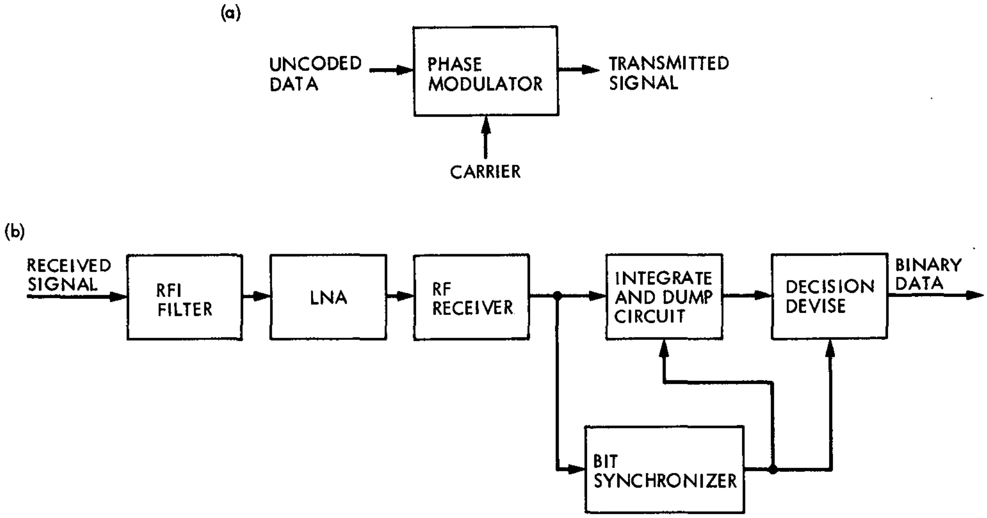


Fig. 1. Uncoded PCM/PM telemetry link block diagram: (a) transmitting system; (b) receiving system

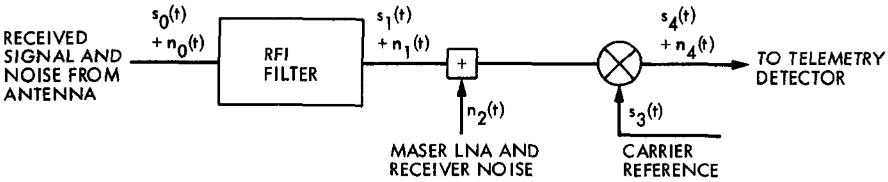


Fig. 2. Analytical model for the RF receiving system

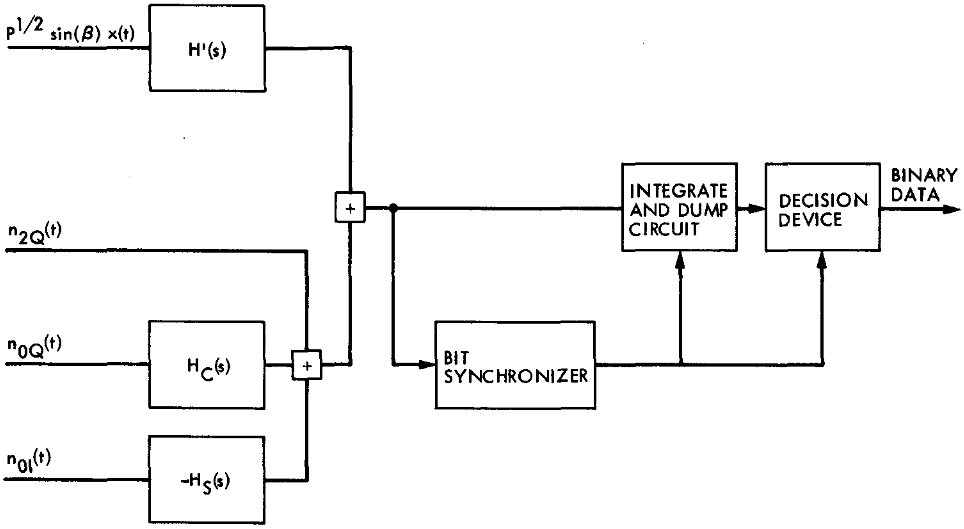


Fig. 3. A simplified analytical model for the DSN RF receiving system and telemetry detector

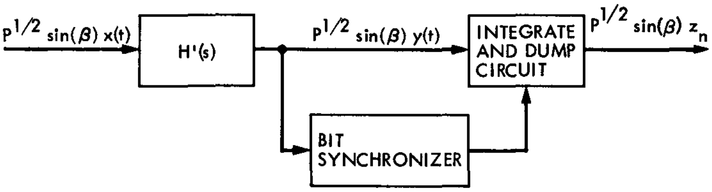


Fig. 4. Analytical model for determining the possible integrate-and-dump circuit output signal levels

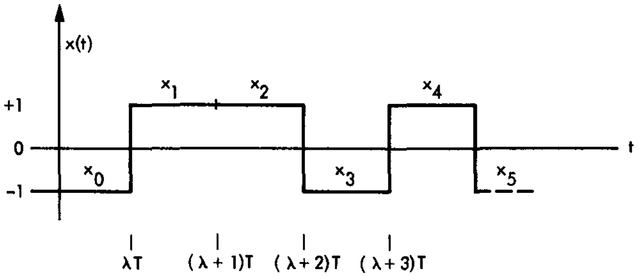


Fig. 5. Sample telemetry data signal x(t)

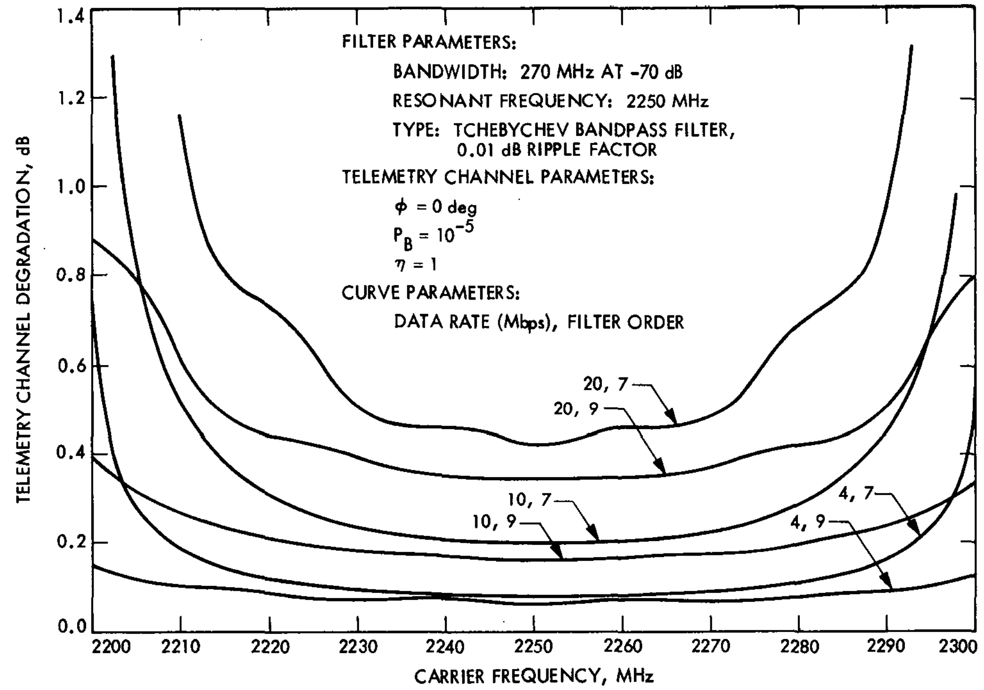


Fig. 6. Telemetry degradation as a function of carrier frequency for the S-band FET amplifier RFI filter assuming all noise originates after the RFI filter

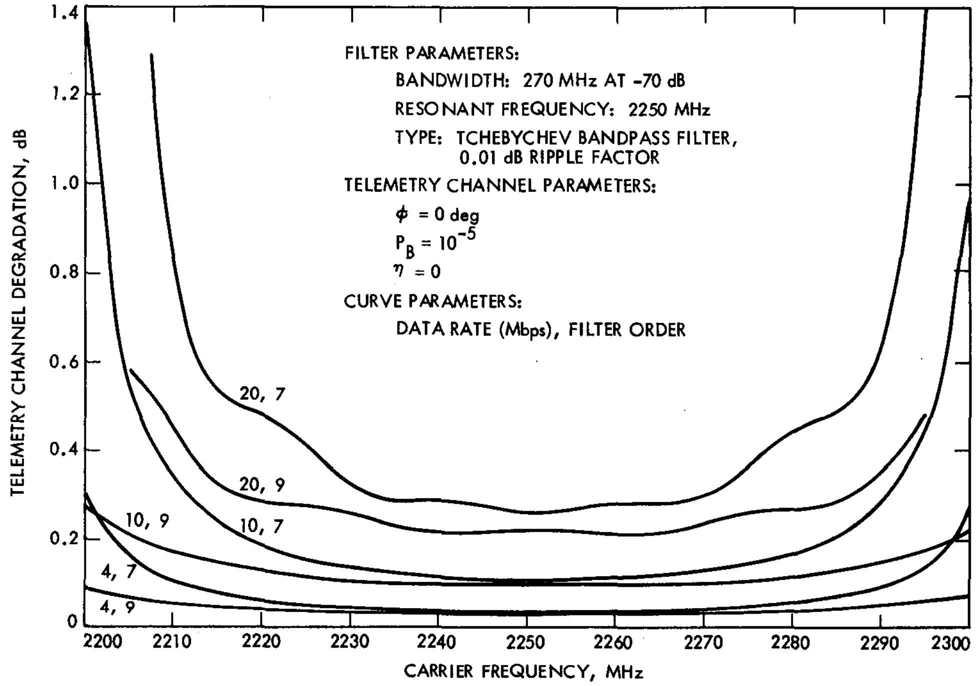


Fig. 7. Telemetry degradation as a function of carrier frequency for the S-band FET amplifier RFI filter assuming all noise originates before the RFI filter

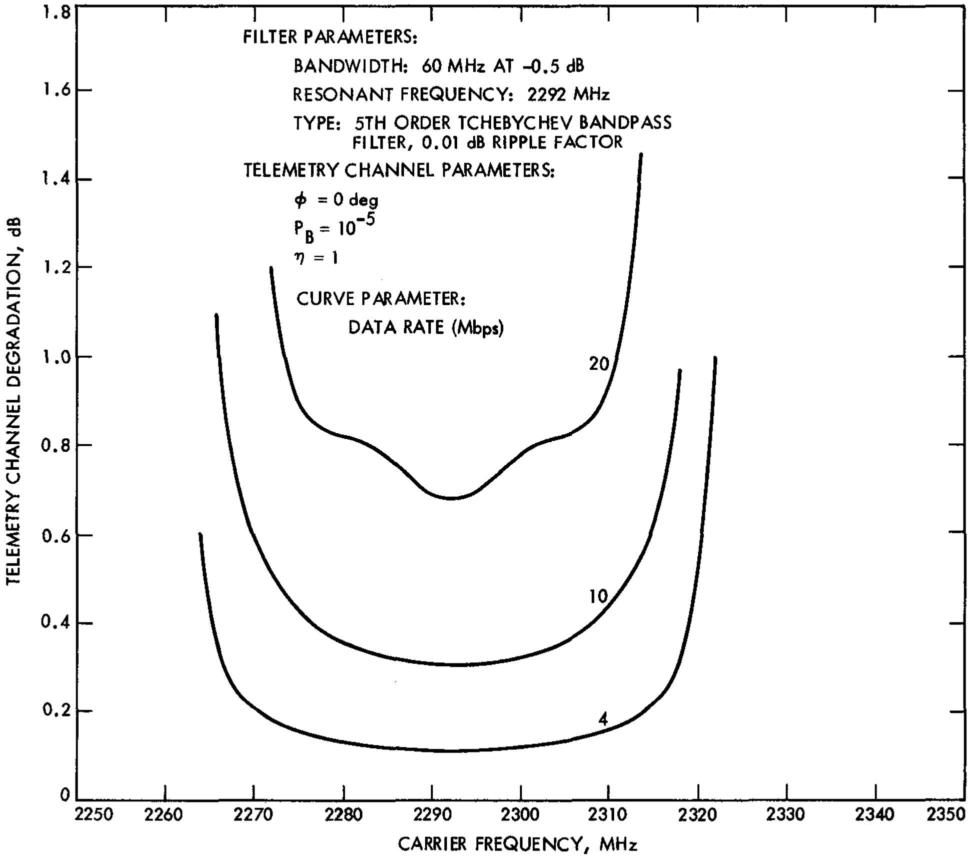


Fig. 8. Telemetry degradation as a function of carrier frequency for the S-band maser amplifier RFI filter assuming all noise originates after the RFI filter

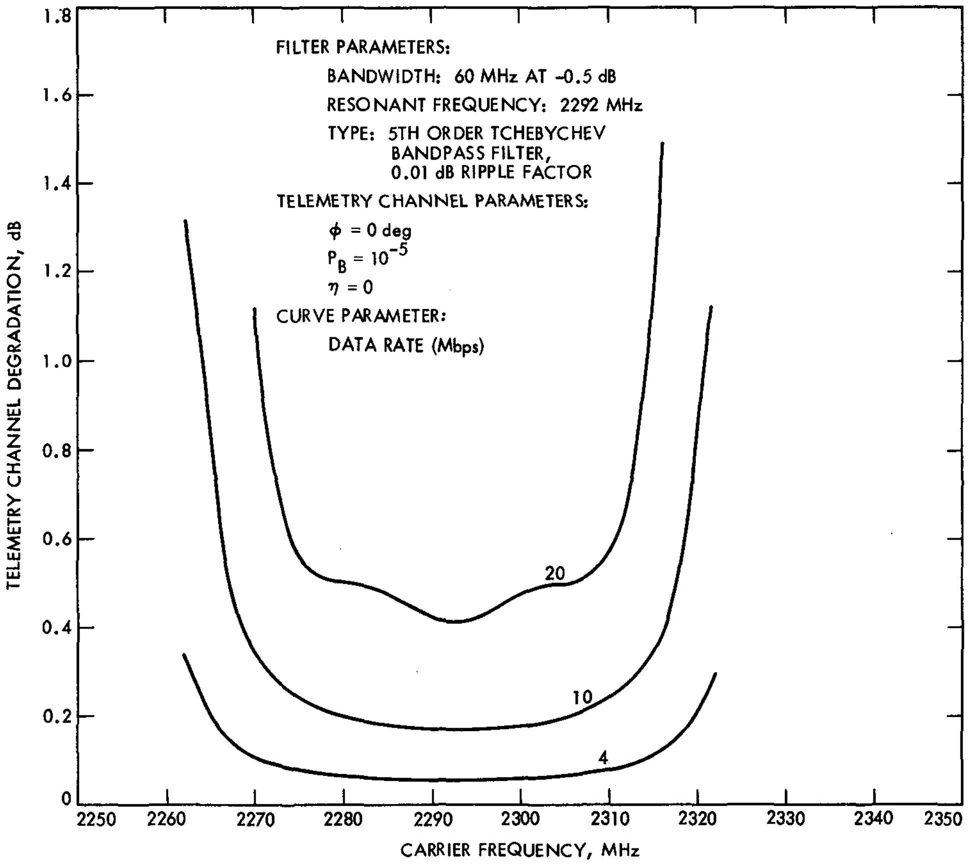


Fig. 9. Telemetry degradation as a function of carrier frequency for the S-band maser amplifier RFI filter assuming all noise originates before the RFI filter

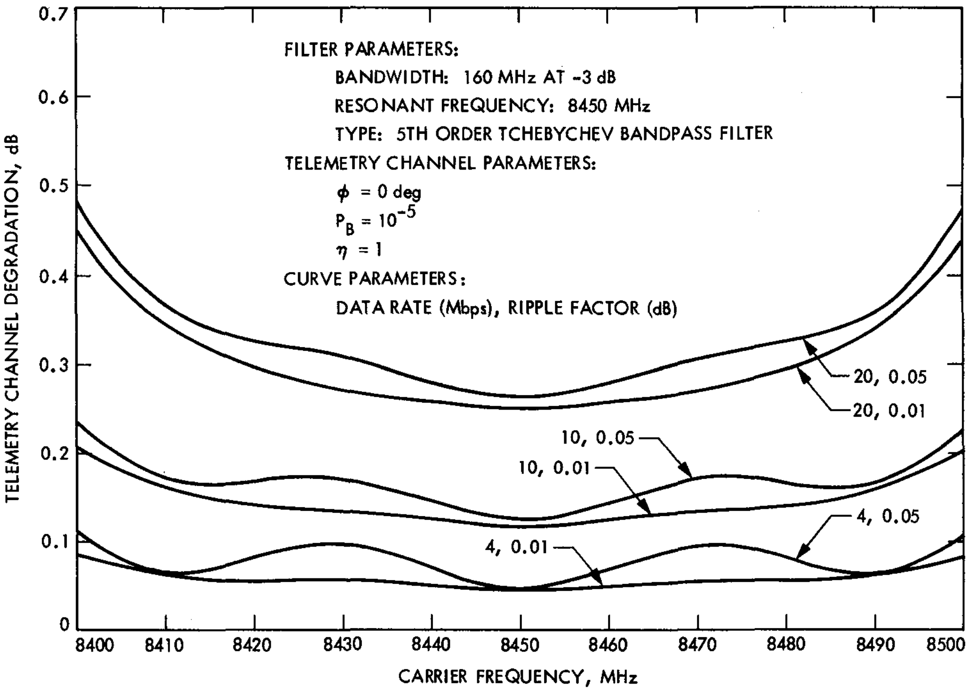


Fig. 10. Telemetry degradation as a function of carrier frequency for the X-band maser amplifier RFI filter assuming all noise originates after the RFI filter

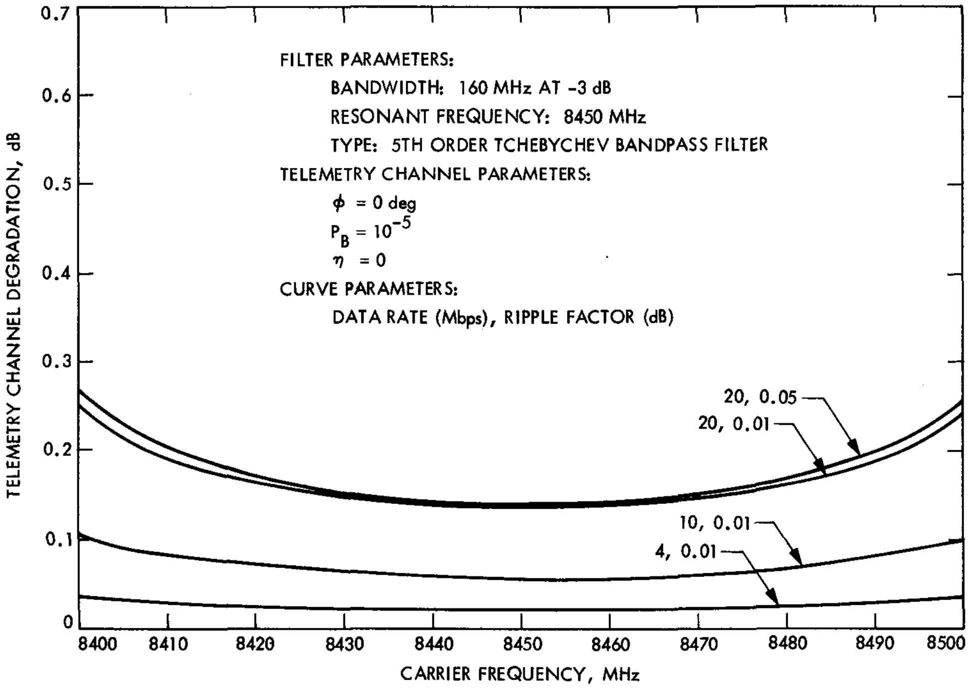


Fig. 11. Telemetry degradation as a function of carrier frequency for the X-band maser amplifier RFI filter assuming all noise originates before the RFI filter

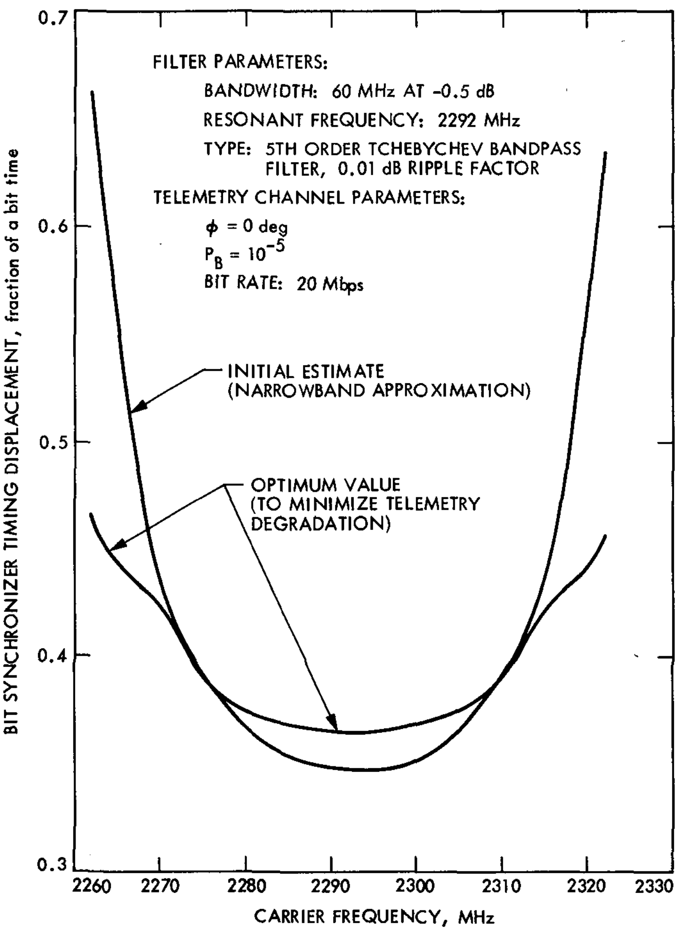


Fig. 12. Comparison of the initial estimate and optimum value of bit synchronizer timing displacement as a function of carrier frequency for the S-band maser amplifier RFI filter

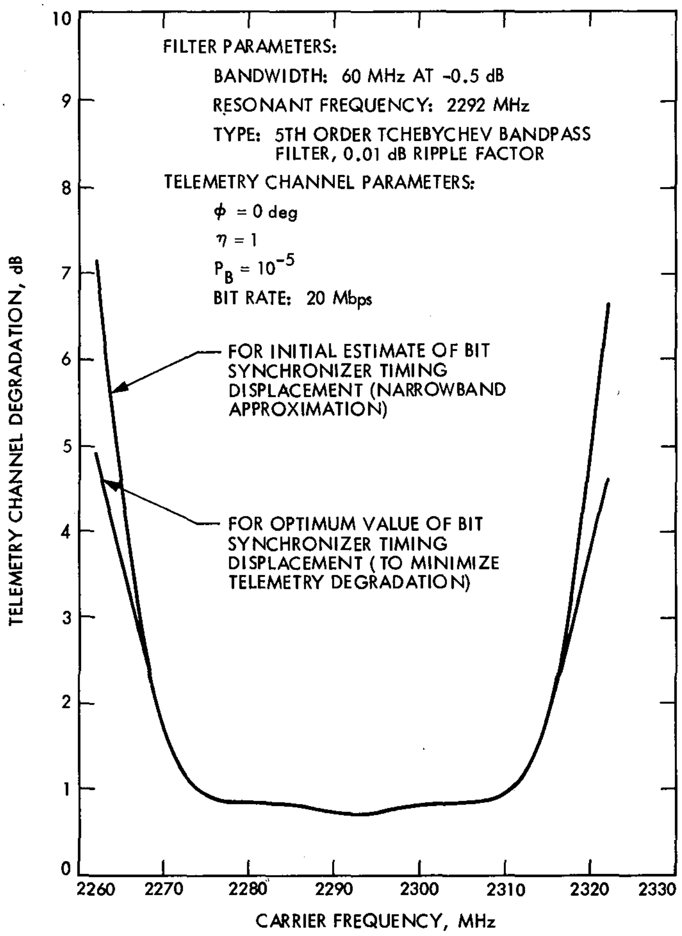


Fig. 13. Comparison of telemetry degradation as a function of carrier frequency for the initial estimate and optimum value of bit synchronizer timing displacement for the S-band maser amplifier RFI filter

Time Interval Errors of a Flicker-Noise Generator

C. A. Greenhall
Communications Systems Research Section

Time interval error (TIE) is the error of a clock at time t after it has been synchronized and syntonized at time zero. Previous simulations of Flicker FM noise have yielded a mean-square TIE proportional to t^2. This study shows that the order of growth is actually $t^2 \log t$, explains the earlier t^2 result, and gives a modified version of the Barnes-Jarvis simulation algorithm.

I. Introduction

Let the time base of a clock be an oscillator whose output is $\sin(2\pi\nu_0 t + \phi(t))$, where ν_0 is a nominal frequency and $\phi(t)$ the phase. Then we call $x_m(t) = \phi(t)/(2\pi\nu_0)$ the measured raw time deviation of the clock. Its derivative $y(t) = dx_m/dt$ is the instantaneous fractional frequency deviation. The τ-average fractional frequency deviation is defined by

$$\bar{y}(t,\tau) = \frac{1}{\tau}\int_{t-\tau}^{t} y(s)ds = \frac{x_m(t)-x_m(t-\tau)}{\tau} \tag{1}$$

The standard measure of frequency stability is the two-sample (Allan) variance $\sigma_y^2(\tau)$, defined by

$$\sigma_y^2(\tau) = \lim_{m\to\infty} \frac{1}{2(m-1)} \sum_{j=2}^{m} [\bar{y}(j\tau,\tau) - \bar{y}((j-1)\tau,\tau)]^2 \tag{2}$$

if the limit exists. For most models of clock noise, the limit does exist and can be computed by an ensemble average, so that

$$\sigma_y^2(\tau) = \frac{1}{2}E[\bar{y}(t,\tau) - \bar{y}(t-\tau,\tau)]^2$$

in which E is mathematical expectation and t is arbitrary. Thus, the two-sample variance gives the mean-square change in τ-average frequency during a time interval of length τ.

On the other hand, the time interval error (TIE), discussed by Kartaschoff (Refs. 1,2), directly measures the error of the clock after it has been synchronized and syntonized. Let the clock be calibrated at time zero, that is, we assume that $x_m(0)$ and Y_0 are available, where Y_0 is some estimate of initial frequency. We could take $Y_0 = y(0)$ if the latter were observable, or $Y_0 = \bar{y}(0,\tau)$ for some τ. The TIE at a later time t is defined by

$$x(t) = x_m(t) - x_m(0) - Y_0 t \tag{3}$$

Since $y = dx_m/dt$, we have

$$x(t) = \int_0^t (y(s) - Y_0)ds \tag{4}$$

The ensemble RMS average of $x(t)$ is called $\sigma_x(t)$ (Ref. 1); thus, by definition,

$$\sigma_x^2(t) = E\,x^2(t) \tag{5}$$

One must keep in mind that $\sigma_x^2(t)$ depends on the choice of Y_0. Also, it is stipulated that Y_0 must not vary with t.

Kartaschoff (Ref. 1) investigated the behavior of TIE for the three random frequency-noise models whose names and properties are given below:

Name	$S_y(f)$	$\sigma_y^2(\tau)$
White FM	h_0	$\frac{h_0}{2\tau}$
Flicker FM	h_{-1}/f	$h_{-1} \ln 4$
Random Walk FM	h_{-2}/f^2	$h_{-2} \frac{2\pi^2}{3} \tau$

where $S_y(f)$ is the one-sided spectral density of the process y, and h_α is a constant. Having conjectured the approximation

$$\sigma_x^2(t) = C t^2 \sigma_y^2(t) \tag{6}$$

for some constant C, Kartaschoff verified Eq. (6) with $C = 1$ by a discrete-time computer simulation of the three FM processes.

On the other hand, a recently-developed structure theory for clock noise processes (Ref. 3) yields the following exact formulas, whose derivation is given in the Appendix:

White FM, $Y_0 = 0$.

$$\sigma_x^2(t) = \frac{h_0}{2} t \tag{7}$$

Flicker FM, $Y_0 = \bar{y}(0,\tau_1)$.

$$\sigma_x^2(t) = h_{-1} t^2 \left(1 + \frac{\tau_1}{t}\right) \left[\ln \frac{t}{\tau_1} + \left(1 + \frac{t}{\tau_1}\right) \ln \left(1 + \frac{\tau_1}{t}\right)\right] \tag{8}$$

$$= h_{-1} t^2 \ln \left(\frac{et}{\tau_1}\right) \left[1 + O\left(\frac{\tau_1}{t}\right)\right] \quad \text{as } t/\tau_1 \to \infty \tag{9}$$

Random Walk FM, $Y_0 = y(0)$.

$$\sigma_x^2(t) = h_{-2} \frac{2\pi^2}{3} t^3 \tag{10}$$

Thus, provided that the calibration frequency Y_0 is properly defined, Eq. (6) holds exactly for White and Random Walk FM with $C = 1$. For Flicker FM, however, the theoretical result Eq. (8) differs from Kartaschoff's simulation result

$$\sigma_x^2(t) = (h_{-1} \ln 4) t^2 \tag{11}$$

by a factor that grows logarithmically with time.

Our purpose here is to explain the discrepancy between Eqs. (8) and (11). Kartaschoff used a flicker-noise generation algorithm of Barnes and Jarvis (Ref. 4) for his simulations. A study of this generator, which is a certain linear filter acting on white noise, shows that the discrepancy is caused, not by any design defect of the filter, nor by any defect of theory, but by the procedure for initializing the generator. In our results, the initialization used by Barnes-Jarvis and Kartaschoff (setting all variables to zero) yields Eq. (11) (although with a different constant factor), while a more complex procedure given in Section IV (making the output process stationary) yields an asymptotic result like Eq. (9), with e replaced by a different constant.

In contrast, the initialization has practically no effect on the observed two-sample variance. No matter which of the two initializations is used, one observes

$$\sigma_y^2(\tau) = h_{-1} \ln 4$$

for τ greater than three times the sample time.

II. The Barnes-Jarvis Generator

The discrete-time n-stage Barnes-Jarvis flicker-noise generator is a cascade of n first-order filters with transfer functions

$$G_j(z) = \frac{z - (1-3\gamma_j)}{z - (1-\gamma_j)}, \quad j = 1,2,\ldots,n \tag{12}$$

where

$$\gamma_j = \frac{1}{6 \cdot 9^{j-1}}$$

For $n = 5$, the frequency response $|H(e^{i\omega})|^2$ of the overall transfer function

$$H(z) = G_1(z) \ldots G_n(z) \tag{13}$$

differs from the ideal two-sided spectrum $h_{-1}\pi/\omega$ by at most 0.25 dB over four decades of frequency, where $h_{-1} = 0.2757$.

The difference equations that implement the generator are

$$y_j(t+1) = (1-\gamma_j)y_j(t) + y_{j-1}(t+1) - (1-3\gamma_j)y_{j-1}(t), \quad j = 1,2, \dots, n \tag{14}$$

where $y_0(t)$ is the input, consisting of "standard" (mean zero and variance one) white noise, and $y_n(t)$ is the output. This implementation differs from that of Barnes-Jarvis and Kartaschoff only in the overall gain and in the order of application of the filter modules.

The system (14) implies that y_n and y_0 are related by an nth-order difference equation whose coefficients can be obtained by expanding the numerator and denominator of $H(z)$ [Eq. (13)] in powers of z. Let the sequence $\psi(0)$, $\psi(1)$, $\psi(2), \dots$ be the impulse response of the overall filter H; thus

$$H(z) = \sum_{t=0}^{\infty} \psi(t)z^{-t}, \qquad |z| \geqslant 1 \tag{15}$$

Assume that the standard white noise $y_0(t)$ is available for all integer t, positive and negative. The unique *stationary* solution $y(t) = y_n(t)$ of the nth order equation is

$$y(t) = \sum_{s=-\infty}^{t} \psi(t-s)y_0(s), \quad -\infty < t < \infty \tag{16}$$

in which $y(t)$ depends on $y_0(s)$ for $-\infty < s \leqslant t$. In the terminology of Box and Jenkins (Ref. 5), y is an ARMA (n,n) process. For a practical simulation, it can be assumed that $y_0(s)$ is available only for $s \geqslant 1$. Thus, it is natural to split $y(t)$ into two independent parts,

$$y(t) = y_+(t) + y_-(t) \tag{17}$$

where

$$\begin{aligned} y_+(t) &= \sum_{s=1}^{t} \psi(t-s)y_0(s), && t \geqslant 1 \\ &= 0, && t \leqslant 0 \end{aligned} \tag{18}$$

is called the *present* part of $y(t)$ because it depends only on the random shocks y_0 after time zero, while

$$\begin{aligned} y_-(t) &= \sum_{s=-\infty}^{0} \psi(t-s)y_0(s), && t \geqslant 1 \\ &= y(t), && t \leqslant 0 \end{aligned} \tag{19}$$

is called the *past* part of $y(t)$ because it depends only on the random shocks before time one. In another terminology, $y_-(t)$ is just the mean-square best linear prediction of $y(t)$ from the past of y_0, while $y_+(t)$ is the prediction error. The stationary process $y(t)$ will also be called the *complete* Barnes-Jarvis (BJ) process.

The time interval error $x(t)$ of the discrete-time frequency process $y(t)$, defined by

$$\begin{aligned} x(t) &= \sum_{s=1}^{t} [y(s) - y(0)], && t \geqslant 1 \\ x(0) &= 0 \end{aligned} \tag{20}$$

also breaks up into present and past parts $x_+(t)$ and $x_-(t)$, obtained from Eq. (20) with y replaced by y_+ and y_-.

The BJ generator, Eq. (14), is initialized by assigning values to $y_1(0), \dots, y_n(0)$. The usual way of doing this, called the *zero* initialization here, is to set all these values, plus $y_0(0)$, to zero (Ref. 4). By an induction on n, one can prove the following assertion: *If the zero initialization is used, then the output $y_n(t)$ of the BJ generator is exactly $y_+(t)$ for $t \geqslant 1$.* Moreover, since $y_+(0) = 0$, $x_+(t)$ is exactly the TIE simulated by Kartaschoff.

Two questions arise. First, are there any practical differences between $y(t)$ and $y_+(t)$? Second, inasmuch as the infinite past of $y_0(t)$ is unavailable, can the complete process $y(t)$ be simulated accurately? Since, for $t \geqslant 1$, both $y(t)$ and $y_+(t)$ (alias $y_n(t)$) satisfy Eq. (14) with identical inputs $y_0(t)$, the difference $y(t) - y_+(t)$ is merely a transient of the filter H; consequently,

$$y(t) - y_+(t) \to 0 \qquad \text{as } t \to \infty \tag{21}$$

Nevertheless, the stationary output $y(t)$ behaves like true nonstationary $1/f$ noise only over a time span of the order of magnitude $6 \cdot 9^{n-1}$, the longest time constant found in H. Over the useful time span of the generator, one cannot expect the transient to be small. As we shall soon see, its effect causes the variances of $x(t)$ and $x_+(t)$ to differ greatly.

The second question also has a yes answer. To simulate the complete BJ process $y(t)$, one has only to set up the initial vector $y_0(0), y_1(0), \ldots, y_n(0)$ with *random* values having the correct joint distribution. The difference equations [Eq. (14)] do the rest. The *stationary* initialization algorithm is given in Section IV.

III. Statistical Results

The mean-square time interval error $\sigma_x^2(t)$ and the two-sample variance $\sigma_y^2(\tau)$ were calculated theoretically for the complete five-stage Barnes-Jarvis process $y(t)$ and for its present part $y_+(t)$. Then, the same quantities were estimated by computer simulation, the zero initialization being used to generate $y_+(t)$ and the stationary initialization (Section IV) being used to generate $y(t)$. Gaussian pseudorandom numbers were used for the white input noise $y_0(t)$. The results are given in Figs. 1 and 2. Figure 1 plots $\sigma_x^2(t)/t^2$ (linear scale) against t/τ_1 (logarithmic scale), where τ_1 is the sample time of the simulation. The mean-square TIE of the complete process dominates that of the present part. To put it another way, *as time goes on, more and more of the TIE (on the average) comes from the remote past.* The results can be summarized by empirical formulas. For the complete process,

$$\sigma_x^2(t) = h_{-1} t^2 \ln(5.5\, t/\tau_1) \tag{22}$$

and for the present part,

$$\sigma_x^2(t) = 2h_{-1}\, t^2 \tag{23}$$

valid for $16 \leqslant t/\tau_1 \leqslant 16000$. Equations (22) and (9) have different constants inside the logarithm. This discrepancy is caused by high-frequency spectral deviation of the Barnes-Jarvis process from pure sampled flicker FM. Equation (23) agrees with Kartaschoff's result Eq. (11), except for the constant factor. This discrepancy is perhaps caused by statistical variations, since Kartaschoff generated only 100 sample functions.

The mean-square TIE $\sigma_x^2(t)$ of the present part can also be interpreted as the minimal mean-square prediction error of the *complete* $x(t)$ from the history of the process before time one. Our result, Eq. (23), is in approximate agreement with a result of Percival for this prediction error variance (Fig. 3.1 in Ref. 6) although our factor $2h_{-1}$ appears to differ slightly from his.

Figure 2 plots $\sigma_y^2(\tau)$ (*not* $\sigma_y(\tau)$) on a linear scale against τ/τ_1 on a logarithmic scale. Here, the situation is quite different. It is apparent that *the two-sample variances of the complete BJ process and of its present part are practically indistinguishable.* In fact, the *past* part of the BJ process, which accounts for most of the TIE for large t, contributes only 9 percent of the complete $\sigma_y^2(4096\tau_1)$ and 0.00015 percent of the complete $\sigma_y^2(\tau_1)$. This explains why earlier BJ users, who relied on a constant $\sigma_y^2(\tau)$ as an indicator of the success of the simulations, were not aware that a large part y_- of the process y was missing.

The error bars in Fig. 2 concern a side issue addressed also by Kartaschoff (Ref. 1), namely, the calculation of the variance of the classical $\sigma_y^2(\tau)$ estimator $S_y^2(\tau,m)$, which is just the right side of Eq. (2) without the limit. The solid bars are the ± one-standard-deviation error bars of $S_y^2(\tau,m)$ as computed for pure Gaussian Flicker FM by Yoshimura (Ref. 7), and later by the author (Ref. 3); the dashed bars come from the sample variances observed during the 2048 simulation runs. The agreement is satisfactory.

IV. Simulation of the Complete BJ Process

It is desired to simulate the stationary solution $y(t) = y_n(t)$ of the Barnes-Jarvis system, Eq. (14). To this end, consider the stationary n-vector process $Z(t) = (Z_1(t), \ldots, Z_n(t))$, $-\infty < t < \infty$, where $Z_j(t) = y_j(t) - y_{j-1}(t)$. (The Z_j are used instead of the y_j because the Z_j are less highly correlated.) The jth component Z_j is obtained by acting on y_0 with the filter whose transfer function is

$$K_j(z) = G_1(z) \ldots G_{j-1}(z)[G_j(z) - 1] \tag{24}$$

[Recall Eqs. (12), (13).] Each $Z_j(t)$ is orthogonal to $y_0(t)$, and the covariance matrix R_Z is given by

$$R_Z(i,j) = E[Z_i(t)Z_j(t)] = \oint_{|z|=1} K_i(z)K_j(1/z)\frac{dz}{2\pi iz} \tag{25}$$

which can easily be evaluated by residues.

Suppose now that we create new random variables $y_0(0)$, $Z_1(0), \ldots, Z_n(0)$ with the above covariances, set $y_j(0) = y_{j-1}(0) + Z_j(0)$, $j=1, \ldots, n$, and generate $y_n(t)$, $t \geqslant 1$, by Eq. (14) from the white noise inputs $y_0(t)$, $t \geqslant 1$. One can then prove that these new $y_n(t)$ have the same covariance structure as the original stationary process.

To set up the $Z_j(0)$, one can use the Choleski factorization of R_Z, namely, $R_Z = LL^T$, where L is an $n \times n$ lower-triangular matrix. The coefficients L_{ij} are given in Table 1 for $n = 6$; for smaller n, one simply truncates the matrix. These coefficients give the $Z_j(0)$ as linear combinations of a standard white noise vector $u_1, \ldots, u_n$. If the simulated process is to be Gaussian, then it is important to make the u_j Gaussian because there are so few of them and their effect is large. Although one could then save time by using uniformly distributed noise inputs $y_0(t)$ (with mean zero, variance one), our simulations used Gaussian noise inputs throughout.

Here, then, is the simulation algorithm for the complete Barnes-Jarvis process. Let GRAN be the routine that generates independent Gaussian random numbers with mean zero, variance one.

```
1. Generate $y_0(0), u_1, \ldots, u_n$ with GRAN
2. For $i = 1$ to $n$
        Let $Z_i = \sum_{j=1}^{i} L_{ij} u_j$   ($L_{ij}$ from Table 1)
        Let $y_i(0) = y_{i-1}(0) + Z_i$
        Next $i$
3. For $t = 0, 1, 2, \ldots$
        Generate $y_0(t+1)$ with GRAN
        For $j = 1$ to $n$
            Compute $y_j(t+1)$ by Eq. (14)
            Next $j$
        Next $t$
```

Remark: To generate the present part of the BJ process, set $y_0(0), y_1(0), \ldots, y_n(0) = 0$, then go to Step 3.

V. Concluding Remarks

As we have seen, the Barnes-Jarvis generator, if given the stationary initialization, produces an output whose mean-square TIE $\sigma_x^2(t)$ has a $t^2 \log t$ behavior, just as theory predicts. The previously-observed t^2 behavior, which appears when the BJ generator is given the zero initialization, can be interpreted as the minimal mean-square error of a linear clock-time predictor based on knowledge of the clock's behavior over the entire past, $-\infty$ to 0. Naturally, this error is smaller than $\sigma_x^2(t)$, the mean-square error of a predictor $x_m(0) + Y_0 t$ for the time deviation $x_m(t)$. We conclude that if the diagnosis of the noise of a clock includes a Flicker FM component, then its $\sigma_x^2(t)$ must include a $t^2 \log t$ component. If $\bar{y}(0,\tau_1)$ is used for calibrating the initial frequency, then this component is given by Eq. (8).

Acknowledgment

The author is pleased to acknowledge a stimulating correspondence with P. Kartaschoff on these subjects.

References

1. Kartaschoff, P., Computer simulation of the conventional clock model, *IEEE Trans. Inst. Meas.*, Vol. IM-28, pp. 193-197, 1979.
2. Kartaschoff, P., Reference clock parameters for digital communications systems, *12th Ann. PTTI Proceedings* (NASA Conference Publication 2175), pp. 515-549, 1980.
3. Greenhall, C., A structure function representation theorem with applications to frequency stability estimation, *IEEE Trans. Inst. Meas.*, Vol. IM-32, pp. 364-370, 1983.
4. Barnes, J., and Jarvis, S., Efficient numerical and analog modeling of flicker noise processes, *NBS Tech. Note 604*, National Bureau of Standards, 1971.
5. Box, G., and Jenkins, G., *Time Series Analysis*, Rev. ed., Holden-Day, San Francisco, 1976.
6. Percival, D., *The statistics of long memory processes*, Ph.D. Thesis, U. of Washington, 1983.
7. Yoshimura, K., Characterization of frequency stability: Uncertainty due to the autocorrelation of the frequency fluctuations, *IEEE Trans. Inst. Meas.*, Vol. IM-27, pp. 1-7, 1978.

Table 1. Coefficients L_{ij} for the stationary initialization of the Barnes-Jarvis process

	j					
i	1	2	3	4	5	6
1	0.603023					
2	0.214635	0.512223				
3	0.301626E-1	0.241088	0.494406			
4	0.345089E-2	0.358003E-1	0.244953	0.491688		
5	0.384698E-3	0.412554E-2	0.366905E-1	0.245520	0.491287	
6	0.427600E-4	0.460283E-3	0.423277E-2	0.368209E-1	0.245599	0.491231

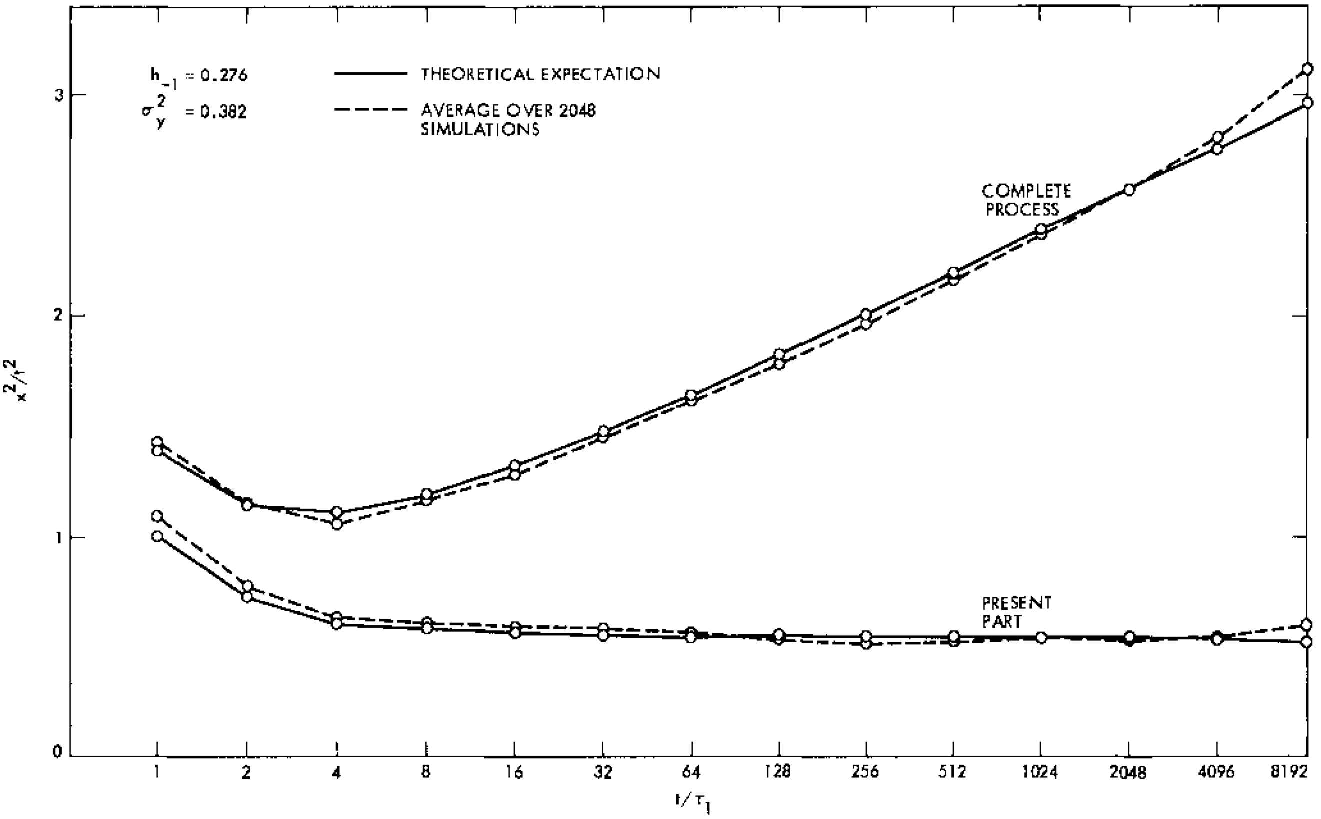


Fig. 1. Squared time interval error of five-stage Barnes-Jarvis Flicker FM model

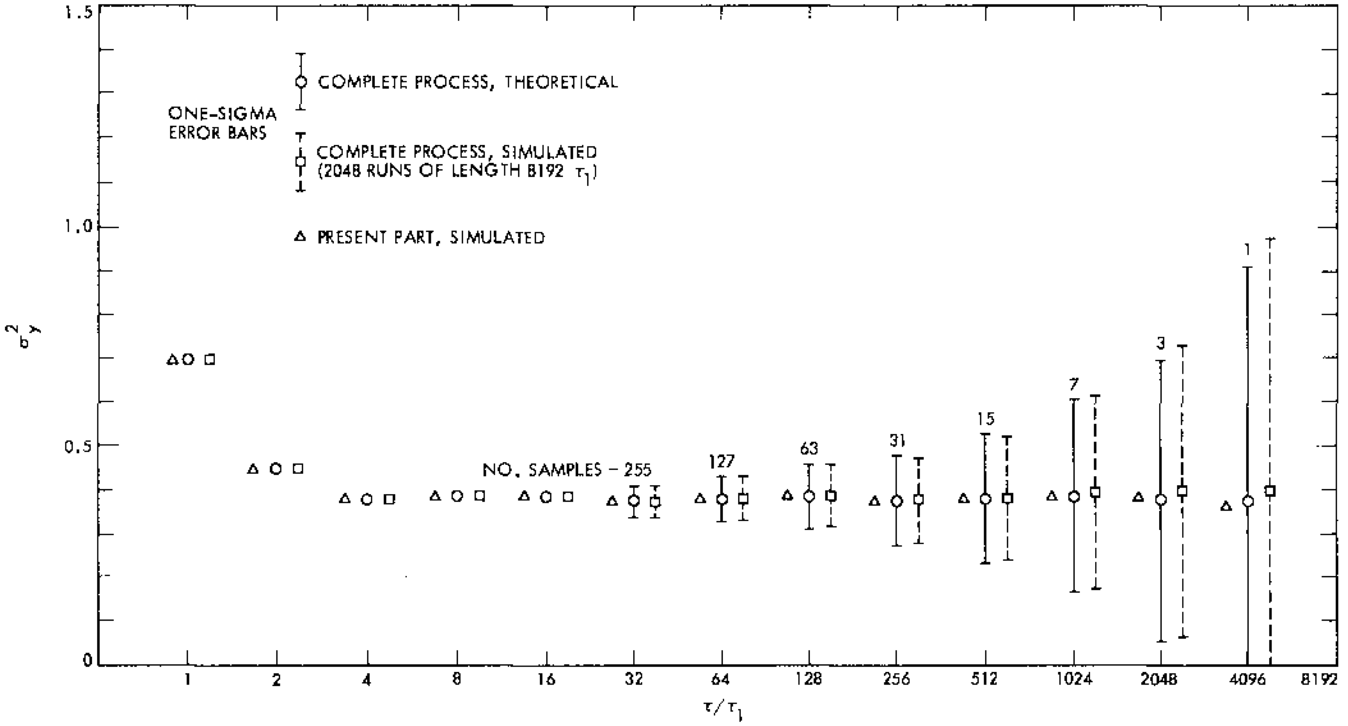


Fig. 2. Two-sample variance of five-stage Barnes-Jarvis Flicker FM model

Appendix

Derivation of Formulas for Mean-Square Tie

Given fixed numbers $t, \tau > 0$, define the second-order difference operator L by its action on a function $f(s)$:

$$Lf(s) = f(s+t) - \left(1 + \frac{t}{\tau}\right) f(s) + \frac{t}{\tau} f(s-\tau) \tag{A-1}$$

where s is used as the time variable because t is fixed. The transfer function of L is

$$L(i\omega) = e^{i\omega t} - \left(1 + \frac{t}{\tau}\right) + \frac{t}{\tau} e^{-i\omega\tau} \tag{A-2}$$

Then, with $Y_0 = \bar{y}(0,\tau)$, Eq. (3) can be written as

$$x(t) = Lx_m(0) \tag{A-3}$$

Now assume that the second differences of the time deviation x_m are stationary. (The examples treated below satisfy this condition.) According to the structure theorem of Ref. 3, there is a function $D(t)$ (appearing as $2\,Re\,C(t)$ in Ref. 3) such that

$$\sigma_x^2(t) = Ex^2(t) = E\,Lx_m(0)\,Lx_m(0) = \Lambda D(0) \tag{A-4}$$

where Λ is the fourth-order operator LL^*, and L^* is the operator with transfer function $L^*(i\omega) = L(i\omega)^*$. Hence the transfer function of Λ is

$$\begin{aligned}\Lambda(i\omega) = |L(i\omega)|^2 = {} & 2\left(1 + \frac{t}{\tau} + \frac{t^2}{\tau^2}\right) - \left(1 + \frac{t}{\tau}\right)\left(e^{i\omega t} + e^{-i\omega t}\right) \\ & - \frac{t}{\tau}\left(1 + \frac{t}{\tau}\right)\left(e^{i\omega t} + e^{-i\omega\tau}\right) \\ & + \frac{t}{\tau}\left(e^{i\omega(t+\tau)} + e^{-i\omega(t+\tau)}\right)\end{aligned} \tag{A-5}$$

Thus, from Eq. (A-4),

$$\begin{aligned}\sigma_x^2(t) = {} & 2\left(1 + \frac{t}{\tau} + \frac{t^2}{\tau^2}\right) D(0) - \left(1 + \frac{t}{\tau}\right)\left(D(t) + D(-t)\right) \\ & - \frac{t}{\tau}\left(1 + \frac{t}{\tau}\right)\left(D(\tau) + D(-\tau)\right) + \frac{t}{\tau}\left(D(t+\tau) + D(-t-\tau)\right)\end{aligned} \tag{A-6}$$

Each noise process has its own D-function to be used in Eq. (A-6).

I. Flicker FM

From Eq. (27) of Ref. 3,

$$D(t) = \frac{h_{-1}}{2} t^2 \ln|t| \tag{A-7}$$

which, when substituted into Eq. (A-6), gives Eq. (8) with $\tau_1 = \tau$.

II. Random Walk FM

Since Λ is a fourth-order operator, one may add a third-degree polynomial to D. Thus, Eq. (28) of Ref. 3 is equivalent to

$$D(t) = \frac{h_{-2}\pi^2}{6} |t|^3 \tag{A-8}$$

and Eq. (A-6) gives

$$\sigma_x^2(t) = h_{-2} \frac{2\pi^2}{3} t^2 (t+\tau) \tag{A-9}$$

As $\tau \to 0$, $\bar{y}(0,\tau) \to y(0)$ because y is a mean-continuous process (Brownian motion), and Eq. (A-9) reduces to Eq. (10). This does not happen for Flicker FM because it has too much power at high frequencies. Consequently, $\bar{y}(0,\tau)$ is well defined, but $y(0)$ is not.

III. White FM

Since $Y_0 = 0$, and x_m has stationary *first* differences, we redefine L as the first-order operator:

$$Lf(s) = f(s+t) - f(s) \tag{A-10}$$

Then everything works as before, except that Λ is now a second-order operator. In place of Eq. (A-6), we have

$$\sigma_x^2(t) = 2D(0) - D(t) - D(-t) \qquad \text{(A-11)}$$

This time, a first-degree polynomial may be added to $D(t)$. Thus, Eq. (26) of Ref. 3 is equivalent to

$$D(t) = -\frac{h_0}{4}\,|t| \qquad \text{(A-12)}$$

which combines with Eq. (A-11) to give Eq. (7), itself a well-known fact.

Minimum-Distance Problems in Protocol Design

E. C. Posner
Office of Telecommunications and Data Acquisition

Z. Reichstein[1]
California Institute of Technology, Student

This article considers codes for use in personal computer file transfer as control characters, when only upper-case ASCII can be used to avoid dependence on unique machine features and promote portability. If ten control functions are needed, a number used in at least one protocol, we seek a subset of ten upper-case ASCII characters with good distance properties. The control functions form themselves naturally into three groups, one of two functions (ACK and NAK) and two of four. We wish to make ACK and NAK as antipodal as possible (distance 6), make the distances within each of the other groups as large as possible (4), and otherwise have as few 2's in the distance table as possible, recognizing that only even distances can occur. We find the minimum and an assignment that attains the minimum. The code is essentially unique. We also solve the analogous problem for two groups of three control functions and one group of four.

I. Introduction

Here we solve a simple coding problem in computer protocol design. People with personal computers want to exchange binary files, but everyone owns a different machine. The problem is that different machines interpret binary seventuples differently. So when setting up communication, it may be impossible to even pass the initial protocol characters to the protocol-handling software. But there is a way out of this. All personal computers agree on what to do with upper-case ASCII. So in this article we restrict protocol control functions to upper-case ASCII. This permits communication to be established between the two computers, after which the files can then be transferred as straight binary files.

Thus, we are in the following situation. We want to choose upper-case ASCII subsets with good distance properties. Before attempting this, we will note that ASCII here is actually 7 bits with even parity adjoined. However, two of the 8 bits are always equal in the upper-case alphabet. This means that we really have a 6-bit code to choose (see Table 1).

Let us formulate this as a precise coding problem. A typical protocol has 10 "block control functions." Here we choose to divide them into three groups, as in Table 2, a "2, 4, 4, Table."

The rationale is that functions within groups need extra protection from each other. This is because within groups, the control functions are more likely to be confusable because the protocol states typically result in outputs that can be or tend to be in only one group. Furthermore, ACK and NAK need to be as unconfusable as possible, to prevent false file transfers.

---

[1] Currently a graduate student at Harvard University

Other groupings are possible, for example a "3, 3, 4 Table" with NAK, ACK, QRY in Group 1, WT, ABH, ABW in Group 2, and EOD, EOT, SOH, SOD in Group 3. We shall consider this case as well.

Let us now formulate this as a coding problem. We are to choose a subset of 10 of the 26 upper-case ASCII odd-parity six-tuples (odd parity, because the invariable 10 has been removed), with the following properties:

a) The distance of ACK and NAK is 6.

b) The minimum distance within Groups 1 and 2 should be as large as possible.

c) There should be as few distances of 2 as possible.

This is for the 2, 4, 4 case. For the 3, 3, 4 case, we want the distances within groups to all be at least 4 (this makes them all 4, it turns out), with as few distances of 2 as possible. We will call a set of three code words of mutual distance 4 an *equilateral triangle* and of four code words of mutual distance 4 a *regular tetrahedron.* We note here that for both problems, it turns out that the restriction to the 26 upper-case ASCII out of the 32 odd-parity six-tuples did not hurt the distance table any, as it turned out in these two cases. Thus we will talk about the alphabet restriction no longer. Probabilistic arguments can be given that make this not too surprising.

II. Intermediate Results

Here we will for convenience revert to even-parity six-tuples instead of odd parity. Call this set E. We shall use lower-case Greek for the elements of E. Note that the Hamming distance d between any two elements of E is even. This section presents five propositions needed in deriving the optimal codes.

We want to solve the following two problems:

Problem 1: Find $\alpha_1, \alpha_2; \beta_1, \beta_2, \beta_3, \beta_4; \gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \mathrm{E}$ such that

$$\left.\begin{array}{lll} d(\alpha_1, \alpha_2) = 6 & & \\ d(\beta_i, \beta_j) \geqslant 4, & i, j = 1, 2, 3, 4, & i \neq j \\ d(\gamma_i, \gamma_j) \geqslant 4, & i, j = 1, 2, 3, 4, & i \neq j \end{array}\right\} \quad (1)$$

and the number of 2's in the distance table of $\alpha_1, \alpha_2; \beta_1, \beta_2, \beta_3, \beta_4; \gamma_1, \gamma_2, \gamma_3, \gamma_4$ is minimal.

Problem 2: Find $\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2, \beta_3; \gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \mathrm{E}$ such that

$$\left.\begin{array}{lll} d(\alpha_i, \alpha_j) \geqslant 4, & i, j = 1, 2, 3, & i \neq j \\ d(\beta_i, \beta_j) \geqslant 4, & i, j = 1, 2, 3, & i \neq j \\ d(\gamma_i, \gamma_j) \geqslant 4, & i, j = 1, 2, 3, 4, & i \neq j \end{array}\right\} \quad (2)$$

and the number of 2's in the distance table of $\alpha_1, \alpha_2, \alpha_3; \beta_1, \beta_2, \beta_3; \gamma_1, \gamma_2, \gamma_3, \gamma_4$ is minimal.

Five propositions will be useful in our search.

Proposition 1: Let $\alpha_1, \alpha_2, \alpha_3 \in \mathrm{E}$ and $d(\alpha_1, \alpha_2), d(\alpha_2, \alpha_3), d(\alpha_1, \alpha_3) \geqslant 4$. Then $d(\alpha_1, \alpha_2) = d(\alpha_2, \alpha_3) = d(\alpha_1, \alpha_3) = 4$.

Proof: Assume the contrary. Say $d(\alpha_1, \alpha_2) > 4$, i.e., $= 6$. Then $\alpha_2^i = \overline{\alpha_1^i}$ where $\overline{x}$ means $1 - x$ for $x = 0, 1$. Thus, for any $a \in \{0, 1\}, |a - \alpha_2^i| = \overline{|a - \alpha_1^i|}$ (i.e., if one of them is 0, then the other is one and vice versa) and, hence, $|a - \alpha_2^i| + |a - \alpha_1^i| = 1$. Then

$$8 \leqslant d(\alpha_1, \alpha_3) + d(\alpha_2, \alpha_3) = \sum_{i=1}^{6} (|\alpha_3^i - \alpha_2^i| + |\alpha_3^i - \alpha_1^i|)$$

$$= \sum_{i=1}^{6} 1 = 6$$

This contradiction proves the proposition.

Proposition 1 shows that we can replace "$\geqslant 4$" by "$= 4$" everywhere in Problems (1) and (2).

Proposition 2: Let $A = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ be a regular tetrahedron. Then each column of the matrix

$$\begin{array}{cccc} \gamma_1^{(1)} & \gamma_1^{(2)} & \cdots & \gamma_1^{(6)} \\ \gamma_2^{(1)} & & \cdots & \gamma_2^{(6)} \\ \vdots & & & \\ \gamma_4^{(1)} & & \cdots & \gamma_4^{(6)} \end{array}$$

has exactly two 0's and two 1's in it, and is unique up to permuting and conjugating (complementing modulo 2) columns.

Proof: We can assume without loss of generality that $\gamma_1 = (0, 0, 0, 0, 0, 0)$; otherwise just conjugate the columns where

$\gamma_i^{(1)} = 1$, i.e., replace 0 by 1 and 1 by 0 in those columns. Then γ_2, γ_3, and γ_4 have exactly two 0's and four 1's. By permuting the columns we can make $\gamma_2 = (1, 1, 1, 1, 0, 0)$. Now γ_3 and γ_4 must have their last two positions be 11. Otherwise their distance from γ_2 would be at most 2. By permuting the first four columns we can make $\gamma_3 = (1, 1, 0, 0, 1, 1)$. Since $d(\gamma_3, \gamma_4) = 4$, γ_4 must be equal to $(0, 0, 1, 1, 1, 1)$.

Thus any regular tetrahedron can be obtained from

$$(0,0,0,0,0,0) = \gamma_1$$

$$(1,1,1,1,0,0) = \gamma_2$$

$$(1,1,0,0,1,1) = \gamma_3$$

$$(0,0,1,1,1,1) = \gamma_4$$

by permuting and conjugating columns.

Since these operations preserve the property we are interested in (exactly 2 zeros in every column), it is sufficient to check it just for the regular tetrahedron above. It, indeed, has exactly two 0's in each column. Hence, any regular tetrahedron also has this property. (Note that we can only conjugate an even number of columns if the tetrahedron is to remain a subset of E.)

Proposition 2 also follows from a more general result on constant-distance codes, but we shall not do it this way.

Proposition 3: Let $B = \{\beta_1, \beta_2, \beta_3\}$ be an equilateral triangle. Then the matrix

$$\begin{matrix} \beta_1^{(1)} & \beta_1^{(2)} & \cdots & \beta_1^{(6)} \\ \beta_2^{(1)} & & \cdots & \beta_2^{(6)} \\ \beta_3^{(1)} & & \cdots & \beta_3^{(6)} \end{matrix}$$

has either one 0 and two 1's or one 1 and two 0's in each column, and is unique up to permuting and conjugating columns.

Proof: We apply the same argument as in Proposition 2. The property we are interested in is again invariant under conjugations and permutations of columns. This operation allows us to transform any equilateral triangle into

$$(0,0,0,0,0,0) = \gamma_1$$

$$(1,1,1,1,0,0) = \gamma_2$$

$$(1,1,0,0,1,1) = \gamma_3$$

as we showed in the proof of Proposition 2. It is easy to see that γ_1, γ_2, and γ_3 have the desired property.

Proposition 4: Let $C = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ be a regular tetrahedron of elements in E, and let γ also be in E. Then

$$\sum_{i=1}^{4} d(\gamma, \gamma_i) = 12$$

Proof: Proposition 2 implies that for $a \in \{0, 1\}$, $k = 1, 2, \ldots, 6$, we have

$$\sum_{i=1}^{4} |a - \gamma_i^{(k)}| = 2$$

for each column k. Thus

$$\sum_{i=1}^{4} |\gamma^{(k)} - \gamma_i^{(k)}| = 2, \qquad 1 \leqslant k \leqslant 6$$

Adding these six equalities together, we get, as stated,

$$12 = \sum_{k=1}^{6} \left(\sum_{i=1}^{4} |\gamma^{(k)} - \gamma_i^{(k)}| \right)$$

$$= \sum_{i=1}^{4} \left(\sum_{k=1}^{6} |\gamma^{(k)} - \gamma_i^{(k)}| \right)$$

$$= \sum_{i=1}^{4} d(\gamma, \gamma_i)$$

Proposition 5. Let $A = \{\alpha_1, \alpha_2, \alpha_3\}$, $B = \{\beta_1, \beta_2, \beta_3\}$ be two equilateral triangles of elements of E. Then

$$\sum_{i,j=1,2,3} d(\alpha_i, \beta_j) \leqslant 30$$

If the sum is 30, then each

$$\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}| = 5$$

for $1 \leqslant k \leqslant 6$ (used for exhaustive search).

Proof: First we show that

$$\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}| \leqslant 5$$

for each $k = 1, 2, \ldots, 6$. The sum

$$\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}|$$

does not change when we simultaneously conjugate all the kth components

$$\alpha_1^{(k)}, \alpha_2^{(k)}, \alpha_3^{(k)}, \beta_1^{(k)}, \beta_2^{(k)}, \beta_3^{(k)}$$

and permute

$$\{\alpha_1^{(k)}, \alpha_2^{(k)}, \alpha_3^{(k)}\}$$

or

$$\{\beta_1^{(k)}, \beta_2^{(k)}, \beta_3^{(k)}\}$$

for each $k = 1, 2, \ldots, 6$.

Thus by Proposition 3 we can assume that $\alpha_1^{(k)}, \alpha_2^{(k)} = 0$ and $\alpha_3^{(k)} = 1$. Then (again by Proposition 3) one of the $\beta_i^{(k)}$ ($i = 1, 2, 3$) is 0 and one of them is 1. Number the 0 first (i.e., $\beta_1^{(k)} = 0$) and the 1 last ($\beta_3^{(k)} = 0$). Now we only have two choices: $\beta_2^{(k)} = 0$ and $\beta_2^{(k)} = 1$. In the first case

$$\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}| = 4$$

and in the second case

$$\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}| = 5$$

Thus

$$\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}| \leqslant 5$$

for each k from 1 to 6. Adding these six inequalities together, we get, as stated,

$$\begin{aligned} 30 &\geqslant \sum_{k=1}^{6} \left(\sum_{i,j=1,2,3} |\alpha_i^{(k)} - \beta_j^{(k)}| \right) \\ &= \sum_{i,j=1,2,3} \left(\sum_{k=1}^{6} |\alpha_i^{(k)} - \beta_j^{(k)}| \right) \\ &= \sum d(\alpha_i, \beta_j) \end{aligned}$$

The second part of the proposition also follows.

III. The Optimal Codes

Figure 1 gives the optimal odd-parity 8-bit ASCII upper-case alphabetic code for the 3, 3, 4 case (2) and Fig. 2 an optimal code for the 2, 4, 4 case (1). Why are they optimal, even when we drop the alphabetic restriction and would be willing to allow *any* even-parity six tuples?

Suppose we have a distance table satisfying (2), with equalities instead of inequalities by Proposition 1. Break it up into blocks as shown in Fig. 3. By Proposition 4 each row sum within Blocks II and III must be 12. Since every entry is 2, 4, or 6, this means that each row in Blocks II and III contains at least two 2's. Thus the total number of 2's in Block II is at least 6 and the total number of 2's in Block III is at least 6.

By Proposition 5 the sum of the nine entries in Block I is at most 30. This implies that Block I contains at least three 2's (7 × 4 + 2 × 2 = 32 > 30). Blocks I′, II′, and III′ are just transposes of Blocks I, II, and III, respectively, and, hence, always have the same number of 2's as Blocks I, II, and III. We see that Fig. 1 has the minimal possible number of 2's in each block. This shows that Fig. 1 is a solution for Problem 2. There are thirty 2's in this table.

The same argument (see Fig. 4) shows that any distance table satisfying (1) must have at least four 2's in Block I, four 2's in Block II, and eight 2's in Block III (7 × 2 + 9 × 4 = 50 > 48). This proves that Fig. 2 has the minimal possible number of 2's in each block and is a solution for Problem 1. There are thirty-two 2's in this table.

If 20 control functions are desired instead of 10, we can use *pairs* of upper-case letters. We have found a constant distance 6 code that does this for 20 functions. This will be reported elsewhere in a more general context.

IV. Uniqueness Results

The code of Fig. 2 with its three groups has a unique automorphism crossing group boundaries. In this, O, H exchange with K, L (or Q, V with U, R), with O corresponding to L and H corresponding to K. Similarly, in Fig. 1, exchange X and Z between groups. This forces V to interchange with O and L with K. Also, in each problem, the two groups of equal size can be swapped.

Are our solutions unique up to the obvious operations? The answer is *yes* for the 2, 4, 4 problem. This is easier to prove than to write up, and we will merely assert it.

The answer is *no* for the 3, 3, 4 problem. In fact, a different distance table can even be achieved, although it can be shown that the two triangles are unique up to obvious transformations. Merely interchange U and A in Fig. 1, and observe that Block II (upper right) becomes the inequivalent (because of the column of all 4's) block

$$\begin{matrix} 2 & 4 & 2 & 4 \\ 4 & 2 & 2 & 4 \\ 2 & 2 & 4 & 4 \end{matrix}$$

This is also inequivalent to Block III, so it is a really different distance table. We shall say no more about these uniqueness problems.

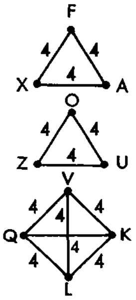
Table 1. Upper case ASCII with even parity

A	0100	0001
B	0100	0010
C	1100	0011
D	0100	0100
E	1100	0101
F	1100	0110
G	0100	0111
H	0100	1000
I	1100	1001
J	1100	1010
K	0100	1011
L	1100	1100
M	0100	1101
N	0100	1110
O	1100	1111
P	0101	0000
Q	1101	0001
R	1101	0010
S	0101	0011
T	1101	0100
U	0101	0101
V	0101	0110
W	1101	0111
X	1101	1000
Y	0101	1001
Z	0101	1010

Table 2. Three groups of block control functions in the 2,4,4 case

Function	Meaning
	Group 0
ACK	Acknowledgement
NAK	Negative acknowledgment
	Group 1
EOD	End of data
EOT	End of text
SOH	Start of header
SOD	Start of data
	Group 2
WT	Wait
QRY	Query
ABH	Abort and hang up
ABW	Abort and wait

	FUNCTION	CHARACTER	8-BIT ASCII
GROUP 0	ACK	F	01000110
	NAK	X	01011000
	WT	A	11000001
GROUP 1	QRY	O	01001111
	ABH	Z	11011010
	ABW	U	11010101
GROUP 2	EOD	V	11010110
	EOT	Q	01010001
	SOH	L	01001100
	SOD	K	11001011



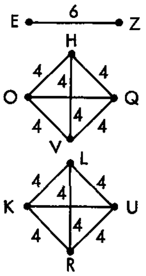
		DISTANCE MATRIX		
		GROUP 0	GROUP 1	GROUP 2
GROUP 0	ACK (F)	0 4 4	2 4 4	2 4 2 4
	NAK (X)	4 0 4	4 2 4	4 2 2 4
	WT (A)	4 4 0	4 4 2	4 2 4 2
GROUP 1	QRY (O)	2 4 4	0 4 4	4 4 2 2
	ABH (Z)	4 2 4	4 0 4	2 4 4 2
	ABW (U)	4 4 2	4 4 0	2 2 4 4
GROUP 2	EOD (V)	2 4 4	4 2 2	0 4 4 4
	EOT (Q)	4 2 2	4 4 2	4 0 4 4
	SOH (L)	2 2 4	2 4 4	4 4 0 4
	SOD (K)	4 4 2	2 2 4	4 4 4 0

3 TWO's PER ROW

3 TWO's

Fig. 1. Optimal code for the 3,3,4 case

	FUNCTION	CHARACTER	8-BIT ASCII
GROUP 0	ACK	E	01000101
	NAK	Z	11011010
GROUP 1	EOD	O	01001111
	EOT	H	11001000
	SOH	Q	01010001
	SOD	V	11010110
GROUP 2	WT	K	11001011
	QRY	L	01001100
	ABH	U	11010101
	ABW	R	01010010



		DISTANCE MATRIX		
		GROUP 0	GROUP 1	GROUP 2
GROUP 0	ACK (E)	0 6	2 4 2 4	4 2 2 4
	NAK (Z)	6 0	4 2 4 2	2 4 4 2
GROUP 1	EOD (O)	2 4	0 4 4 4	2 2 4 4
	EOT (H)	4 2	4 0 4 4	2 2 4 4
	SOH (Q)	2 4	4 4 0 4	4 4 2 2
	SOD (V)	4 2	4 4 4 0	4 4 2 2
GROUP 2	WT (K)	4 2	2 2 4 4	0 4 4 4
	QRY (L)	2 4	2 2 4 4	4 0 4 4
	ABH (U)	2 4	4 4 2 2	4 4 0 4
	ABW (R)	4 2	4 4 2 2	4 4 4 0

4 TWO's PER ROW

3 TWO's PER ROW

4 TWO's

3 TWO's

Fig. 2. Optimal code for the 2,4,4 case

	$\alpha_1 \alpha_2 \alpha_3$	$\beta_1 \beta_2 \beta_3$	$\gamma_1 \gamma_2 \gamma_3 \gamma_4$
α_1 α_2 α_3	**4**	BLOCK I	BLOCK II
β_1 β_2 β_3	BLOCK I'	**4**	BLOCK III
γ_1 γ_2 γ_3 γ_4	BLOCK II'	BLOCK III'	**4**

Fig. 3. Block structure for 3,3,4

	$\alpha_1 \alpha_2$	$\beta_1 \beta_2 \beta_3 \beta_4$	$\gamma_1 \gamma_2 \gamma_3 \gamma_4$
α_1 α_2	**6**	BLOCK I	BLOCK II
β_1 β_2 β_3 β_4	BLOCK I'	**4**	BLOCK III
γ_1 γ_2 γ_3 γ_4	BLOCK II'	BLOCK III'	**4**

Fig. 4. Block structure for 2,4,4

Note on the Optimum Search Strategy for Uniformly Distributed CW Transmitters

S. Gulkis
Atmospheric Sciences Section

The relative probability of detecting randomly distributed CW transmitters as a function of the fraction of the sky which is searched (in a fixed time) is given. It is shown that the probability of detecting such a class of transmitters with a given receiving system is a maximum if the entire sky is searched. The particular case of a search in which the number of directions searched is equal to the telescope gain and the integration time per beam element is equal to the reciprocal of the channel bandwidth is discussed.

I. Introduction

It is important in designing a search program for SETI (Search for Extraterrestrial Intelligence) to understand the factors which effect the probability for detecting signals from distant transmitters. An optimum search strategy cannot be identified at the present time since the statistical properties of the transmitters and the motivations of the senders are not known. Nevertheless, a simple model, in which the transmitters are uniformly distributed in space and have a power law distribution of intrinsic transmitted power, appears to have sufficient generality to provide some guidance in designing a search program. We examine, in this note, the relative probability of detecting signals from such a class of transmitters in a fixed period of time, as a function of the fraction of sky which is searched. The scenario envisaged is that the observer may choose between scanning slowly (and achieving high sensitivity in a few directions) or scanning the entire sky (at reduced sensitivity).

We show that the probability of detecting such a class of CW transmitters is maximized if the entire sky is searched when the received signal-to-noise is low. This means that one is more likely to detect a weak CW signal by scanning the entire sky than by concentrating on a few areas. Since strong CW signals can be detected either by focusing the search in a few areas or by scanning the entire sky, the primary interest in this paper is on the weak signal case.

The detection of pulsed signals is not considered in this note. The detection of these signals may favor one scanning technique over another depending on factors such as the duty cycle of the transmitters.

II. The Probability Relationship

Derivation. Following Drake (Ref. 1), we assume that signals from civilizations radiating an equivalent isotropic radiated power (EIRP), P_o, are originating from uniformly spaced locations having a density of (n_o) transmitters per unit volume. We ignore the fact that the galaxy is a highly flattened disk and consider the spherical volume in the vicinity of the sun where stars are distributed more or less at random except for a tendency to cluster. This volume has a maximum radius of approximately 1 kiloparsec. The maximum range of a given search system is related to the minimum detectable flux, S_m, by the expression

$$R = \left(\frac{P_o}{4\pi S_m}\right)^{1/2} \tag{1}$$

and the total number of detectable signals within the spherical volume is given by

$$N_{\text{det}} = \left(4\pi\frac{n_o}{3}\right)\left(\frac{P_o}{4\pi S_m}\right)^{3/2} \tag{2}$$

The minimum detectable flux, S_m, depends on the diameter, D, of the receiving antenna, the system noise temperature, T_s, and the receiver channel bandwidth, B_s. Oliver and Billingham (Ref. 2) show that the minimum detectable (coherent) flux, defined as being equal to the rms fluctuations due to the noise, is equal to

$$S_m = \frac{kT_sB_s}{\frac{\pi D^2}{4}}\,\frac{1+(1+n)^{1/2}}{n} \tag{3}$$

and the range at which a signal can be detected with a given receiving system is given by

$$R = \left(\frac{D}{4}\right)\left[\frac{P_o}{kT_sB_s}\,\frac{n}{(1+(1+n)^{1/2})}\right]^{1/2} \tag{4}$$

In these expressions

k = Boltzmann's constant

T_s = system noise temperature

D = receiving antenna diameter

B_s = receiver channel bandwidth

t = integration time

$n = (B_st)$ = number of independent samples averaged

Drake (Ref. 1) shows that the probability of success of an observation program is proportional to (a) the total frequency searched, (b) the total solid angle searched, and (c) the spherical volume defined by the radius R. These proportionalities lead to the following expression for the probability of success, P_s of a search over M different directions using a system having C channels (total bandwidth = CB_s), and an antenna beam solid angle equal to Ω_s:

$$P_s = Kn_oMC\left(\frac{B_s}{B_t}\right)\left(\frac{\Omega_s}{\Omega_t}\right)R^3 \tag{5}$$

In this expression, K is a constant of proportionality, Ω_t (= 4π) is the total solid angle of the sky, and B_t is the total bandwidth range in which the signal is confined. The ratio $(M\Omega_s/\Omega_t)$ is the fraction of the sky which is surveyed and the ratio (CB_s/B_t) is the fraction of the total bandwidth which is surveyed.

Noting that the telescope gain, solid angle, the diameter are related through the equations

$$G = \left(\frac{\pi D}{\lambda}\right)^2 = \frac{1}{\Omega_s} \tag{6}$$

and the total observation time, ϕ, number of sky elements examined, M, and the number of independent samples at each sky element, n, are related through the equations

$$n = B_st = B_s\frac{\phi}{M} \tag{7}$$

the probability of success can be written as follows

$$P_s = Kn_oMC\left(\frac{\Omega_s}{\Omega_t}\right)\left(\frac{B_s}{B_t}\right)\left(\frac{D^3}{64}\right)\left[\frac{P_o}{kT_sB_s}\,\frac{n}{1+\sqrt{1+n}}\right]^{3/2} \tag{8}$$

$$= Kn_oMC\frac{\lambda^3}{(4\pi)^4}\left(\frac{B_s}{B_t}\right)G_2^{\;1/2} \times\left[\frac{P_o}{kT_sB_s}\,\frac{B_s\phi}{M+\sqrt{M^2+MB_s\phi}}\right]^{3/2} \tag{9}$$

In the derivation of Eq. (9), we assumed that each transmitter had the same intrinsic power. It is easy to generalize the equation for the case of a continuous distribution of powers by rewritting Eq. (9) as a differential probability of success.

$$dP_s = KMC\frac{\lambda^3}{(4\pi)^4}\left(\frac{B_s}{B_t}\right) \times G^{1/2}\left[\frac{1}{kT_sB_s}\,\frac{B_s\phi}{M+\sqrt{M^2+MB_s\phi}}\right]^{3/2}P^{3/2}\,P(P)dP \tag{10}$$

If the distribution of powers is a power law of the form

$$P(P) = K_1 P^{-\alpha} \tag{11}$$

Equation (10) may be integrated between the limits P_u and P_L to yield the following expression for the probability of success:

$$P_s = KK_1MC \frac{\lambda^3}{(4\pi)^4}\left(\frac{B_s}{B_t}\right) \times G^{1/2}\left[\frac{1}{kT_sB_s}\,\frac{B_s\phi}{M+\sqrt{M^2+MB_s\phi}}\right]^{3/2} \times \frac{P_u^{5/2-\alpha}-P_L^{5/2-\alpha}}{5/2-\alpha} \tag{12}$$

It is seen from this equation that the spectral index, α, determines whether the strong sources or the weak sources will dominate the probability of success. If the spectral index is less than 5/2, then the strong sources, although less numerous will dominate. If the spectral index is greater than 5/2, then the weaker and more numerous sources will dominate. If the spectral index is exactly 5/2, then all sources regardless of their power will contribute uniformly to the probability of success.

In Eqs. (9) and (12), we have implicitly assumed that the telescope gain and solid angle are constant over the receiving bandwidth of the system ($C\,B_s$). This is true in most high performance antenna receiving systems, but may break down if the receiving bandwidth is very large. In this case, an integration over the wavelength needs to be performed.

Interpretation. Figure 1 shows the dependence of the probability on the number of different beam areas examined, M, for a number of different telescopes whose gains vary from 10^2 to 10^8. The value of $B_s\,\phi$ used is 10^8. We note in this figure that for any given telescope, the probability increases as M increases up to the point where the telescope gain becomes equal to the number of directions examined. M cannot exceed the gain since the gain sets the limit on the number of independent areas in the sky which can be examined.

Another constraint on the number of different directions which can be examined is superimposed by the system channel bandwidth, B_s. Since independent measurements cannot be made more frequently than the reciprocal of the bandwidth, the product of the channel bandwidth and total observation time is an upper limit on the number of independent directions in the sky which can be examined in a specified time.

$$M \leqslant B_s\phi \tag{13}$$

This limit is in addition to the constraint set by the number of resolvable directions in the sky. If an attempt is made to increase M beyond this critical product, beam smearing will occur, and it will not be possible to differentiate spatial directions on the scale of the antenna beam size.

Figure 1 shows that it is advantageous to carry out a search by searching the entire sky (M = Gain) with the highest gain (largest) antenna available. This result holds provided that the telescope gain is less than the product $B_s\phi$. Figure 2 shows how the probability of success varies with the telescope gain under the following assumptions:

$$G < B_s\phi \qquad M = G \text{ (entire sky is surveyed)}$$

$$G > B_s\phi \qquad M = B_s\phi\,(M\Omega_s < 4\,\pi) = 10^8$$

Figure 2 shows that the probability of success increases as $(\text{Gain})^{3/4}$ up to the critical point, $B_s\phi = G$; at larger values of gain, the probability increases as $(\text{Gain})^{1/2}$. This figure shows that it is always better to use a large antenna rather than a small antenna. Of course there are mechanical considerations which may favor the use of a smaller antenna; these are not discussed here.

It is of interest to make a comparison between two identical receiving systems which both operate for the same time but with different amounts of the sky being examined. The ratio of the probability of success of a search which examines M different directions, to that of a search which examines the entire sky ($M = G$ = gain) is given by

$$\frac{P_s(M)}{P_s(G)} = \frac{M}{G}\left[\frac{G+\sqrt{G^2+GB_s\phi}}{M+\sqrt{M^2+MB_s\phi}}\right]^{3/2} \tag{14}$$

Figure 3 shows this relationship for the particular case of a 60 dB gain antenna and a channel bandwidth-total observation time product, $B_s\phi = 10^8$. The figure clearly illustrates that the relative probability for detecting a weak signal increases as more directions are searched.

III. The Special Case of the Matched System

It is of interest to consider the special case in which the integration time per beam element is matched to the channel bandwidth, B_s, such that $B_st = 1$, and the telescope gain is matched to the total observation time such that $G = M = B_s\phi$.

These two conditions define a matched system and lead to the following expression for the probability:

$$P = \frac{K^3}{(4\pi)^4} n_o C \left(\frac{B_s}{B_t}\right) \phi^{3/4} \left(\frac{P_o}{kT_s}\right)^{3/2} \qquad (15)$$

It is noticed that this expression is independent of the telescope gain. This at first sight appears to be in contradiction with our previous result which stated that larger antennas produce a higher probability of success than smaller antennas. However, since the matched condition requires that channel bandwidth vary with the telescope gain, the previous result which assumed the bandwidth to be constant does not apply.

In order for matched systems to have the same probability of success, they must cover the same total bandwidth. Since matched low gain antenna systems utilize smaller channel bandwidths than higher gain antenna systems (channel bandwidth is proportional to Gain in matched systems), they must compensate by using a larger number of channels. The table below shows parameters for three matched systems with the same probability of success. The total observation time is assumed to be 10^7 seconds. The highest gain system is assumed to have a channel bandwidth of 1 Hz and a total bandwidth of 1 KHz.

Table 1 illustrates the ultranarrow channel bandwidths and large number of total channels which are required for a small antenna to have the same probability of success as a large system. Since the interstellar medium imposes a minimum usable bandwidth (i.e., Ref. 3), the bandwidth cannot be reduced indefinitely. Hence, in practice, a small antenna system cannot be made as sensitive as a large antenna system, even if the number of channels required imposes no limits.

IV. Discussion

We have shown in this note that the probability of detecting a weak signal from randomly distributed transmitters increases faster by increasing the fraction of sky that is searched than by increasing the sensitivity in a given direction. The reason why it is more advantageous to search a larger solid angle at reduced sensitivity than to search a smaller solid angle at increased sensitivity can be understood through the following argument. The number of detectable transmitters increases as the minimum detectable flux raised to the $-3/2$ power [Eqs. (1) and (2)]. Since the minimum detectable flux decreases as the inverse square root of time for weak signals, the number of detectable transmitters in a given direction increases as the observation time raised to the 3/4 power. On the other hand, the number of detectable transmitters is also proportional to the solid angle searched. Since the solid angle searched can be made to increase in proportion to the time by searching in different directions, the number of detectable transmitters can be made to increase in direct proportion to the time (rather than to the 3/4 power). Hence, the probability of success increases faster by scanning the entire sky than by concentrating in the search to a few directions. The ratio of the number of transmitters which could be detected with a given telescope system in a given time if the entire sky is searched to those that could be detected in the same time if a single direction is observed is proportional to the observation time raised to the 3/4 power.

Although this result was derived for the case of coherent radiation, it applies more generally to detection situations in which the minimum detectable flux decreases more slowly than the observation time raised to the $-2/3$ power. For example, the minimum detectable flux varies as the inverse square root of time for incoherent radiation. Hence the result holds for incoherent as well as (weak) coherent radiation. The result breaks down whenever the minimum detectable flux decreases faster than the observation time raised to the $-2/3$ power. One example where this occurs is in the detection of strong coherent signals. In this case, the flux varies inversely with time.

Aside from mechanical considerations, a large telescope can always be used to produce a higher probability of success than a smaller telescope. Also, it is possible that a small telescope which surveys the entire sky will have a higher probability of success than a larger telescope that concentrates a search in a few directions.

The probability of success of a "matched" system is the same for all telescopes regardless of their gain provided that they cover the same bandwidth. This requires that small telescopes use narrow bandwidths and thus a larger number of channels.

Acknowledgments

I would like to thank F. Drake, M. Janssen, M. Klein, T. Kuiper, A. Lokshin, B. Oliver, E. Olsen, and J. Tarter for reading the manuscript and making helpful suggestions.

References

1. Drake, F. D. (1983), Estimates of the Relative Probability of Success of the SETI Search Program, *SETI Science Working Group Report*, NASA Technical Report, Eds., F. Drake, J. H. Wolfe, C. L. Seeger, October 1983, p. 67.
2. Oliver, B. M., and Billingham, J. *Project Cyclops: A Design Study of a System for Detecting Extraterrestrial Intelligent Life*, prepared under Stanford/NASA/Ames Research Center, 1971 Summer Faculty Fellowship Program in Engineering System Design, NASA Report CR 114445, p. 56.
3. Drake, F. D., and Helou, G. (1977), The Optimum Frequencies for Interstellar Communications as Influenced by Minimum Bandwidths, Report 76, National Astronomy and Ionosphere Center, Cornell University, Ithaca, New York.

Table 1. Channel bandwidths and total of channels required for a small antenna to have the same probability of success as a large system

Gain	Observation Time per Sky Element	Number of Elements	Bandwidth	Number of Channels
10^6	10	10^6	1	10^3
10^5	100	10^5	0.1	10^4
1	10^7	1	10^{-6}	10^9

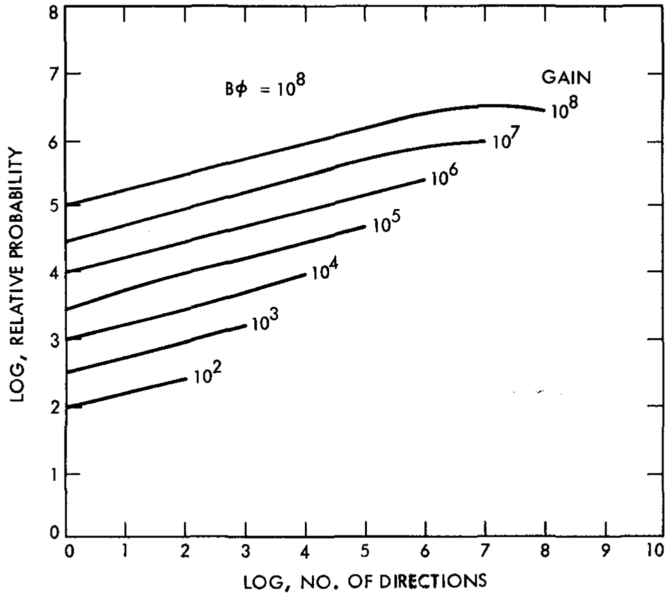


Fig. 1. Relative probability of success for a number of different antennas whose gains vary from 10^2 to 10^8 as a function of the number of directions in the sky which are searched. The search time–channel bandwidth product is taken to 10^8.

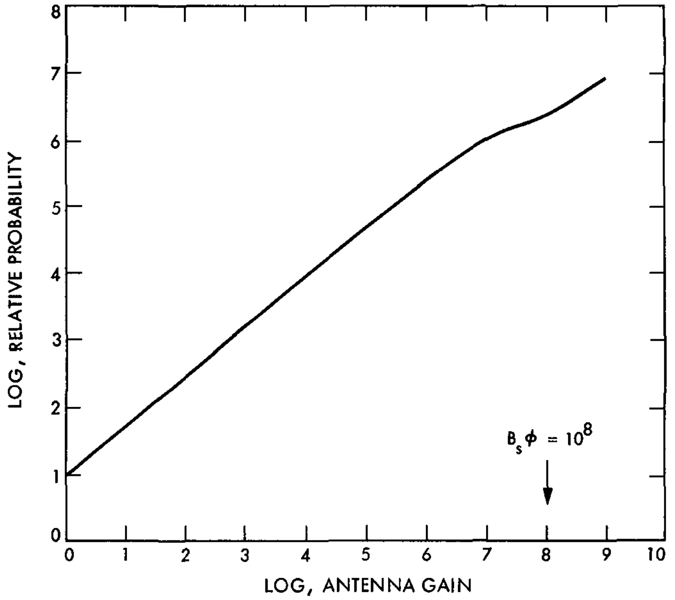


Fig. 2. Relative probability of success as a function of the antenna gain. The search time–channel bandwidth product is taken to be 10^8.

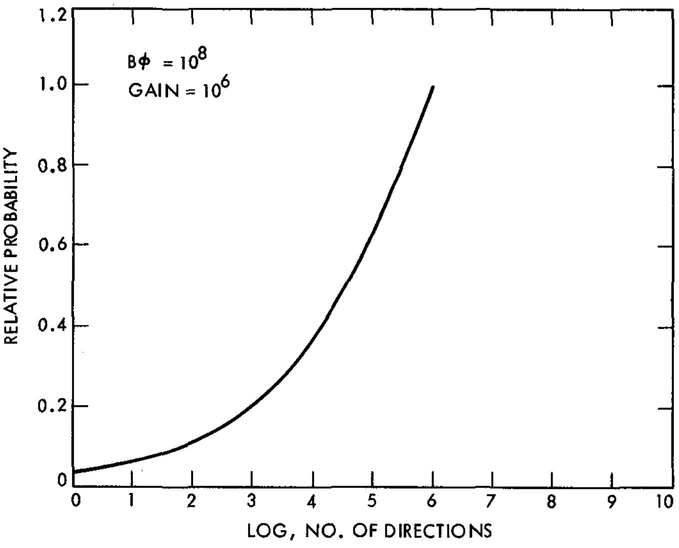


Fig. 3. Relative probability of success as a function of the number of directions in the sky which are searched for a 60 dB antenna. Curve is normalized to unit probability at 10^6 directions which corresponds to the entire sky being searched. The search time–channel bandwidth product is taken to be 10^8.

An Investigation of the Effects of Scan Separation on the Sensitivity of the SETI All Sky Survey for the Case of Gaussian Noise

A. Lokshin and E. T. Olsen
Atmospheric Sciences Section

This article presents an analysis of the scalloping problem for the case of Gaussian noise statistics. We derive the optimal weighting strategy for linearly combining two observations in adjacent beam areas, and compare the sensitivity and scalloping for this weighting strategy with that realized using a single observation or using equal weighting of two observations. We also calculate the variation of the probability for detecting ETI signals with scan separation for the various weighting strategies, assuming that the transmitters are of equal strength and are uniformly distributed throughout space.

I. Introduction

One component of the Search for Extraterrestrial Intelligence (SETI) will be a survey of the entire celestial sphere over a broad frequency range to a significantly low limiting flux (Ref. 1). The inherent advantage of this strategy is that all directions are observed, and thus any signal which exceeds the threshold of the search will be located. The SETI program must design an efficient sky survey strategy which realizes the survey goals within the constraints of available antenna resources.

Time constraints and antenna dynamics dictate that the survey will be carried out by smoothly sweeping the beam across the sky. At the end of each scan, the motion must be reversed without exciting the antenna's natural modes of mechanical oscillation. At the same time the pointing in the orthogonal direction must be stepped so that the subsequent scan traverses a neighboring strip of the celestial sphere. Over a period of time, all of the celestial sphere available to the antenna will have been observed. Unless care is taken in the design of the survey, however, the sensitivity to an ETI signal will be a periodic function of position with respect to the scan pattern. This feature is commonly referred to as scalloping and its magnitude is defined to be the quantity of minimum sensitivity minus maximum sensitivity.

This article is one of a series of technical reports proceeding from the SETI sky survey definition studies and presents an analysis of the scalloping problem for the case of Gaussian noise statistics and ETI sources whose signal strength does not change with time. We derive an optimal weighting strategy for linearly combining data in two adjacent beam areas and then compare the sensitivity and scalloping achieved using this strategy with that realized utilizing practical weighting strategies. Finally, we derive the variation with scan separation of the probability for detecting ETI signals, assuming that the transmitters are of equal strength and uniformly distributed throughout space. Subsequent papers will extend our analyses to the case of non-Gaussian noise statistics, the realm in which the contemplated sky survey will be operating due to the small number of independent samples comprising an accumulation.

II. Derivation of the Optimal Weighting Strategy

In this section, we derive an optimal strategy for linearly combining two observations to achieve the maximum sensitivity to a signal. To simplify the calculation, we assume a one dimensional model (see Fig. 1). The x-axis represents the loci of possible source positions between scan tracks, and we search by stepping a symmetrical beam, $f(x)$, in increments of x_0 between the scans. The y-axis represents the gain of the beam, and the scan direction is normal to the plane of the figure. Suppose that two neighboring beams straddle the source position, x_s, and that the first beam is located on the origin of the x-axis and is the nearer of the two to x_s. Thus the gain of the first beam applied to the signal, f_1, is larger than that of the second beam, f_2.

The criterion for detection of the signal is an excess noise power observed in the receiver attached to the one dimensional antenna. If the beam is located far from the source position, the noise power distribution function is Gaussian with an expectation, $\langle n \rangle$, and a variance, σ^2. If the beam is near the source position, the noise power distribution function is different in that its expectation is augmented by the product of the source strength and beam gain (see Fig. 2). We wish to design an efficient algorithm to detect this signal consistent with a previously set probability that it is caused by noise alone, α_0, and probability of missing the signal, β_0.

In the discussion which follows we shall employ the notation:

f_1 = the gain of the (nearer) first beam area at x_s

f_2 = the gain of the second beam area at x_s

n_1, n_2 = power received due to background noise in each of the two beam areas

s = time invariant strength of the source

p_1 = received power from first beam = $n_1 + f_1\, s$

p_2 = received power from second beam = $n_2 + f_2\, s$

Th = power threshold which must be exceeded to satisfy detection criterion

α_0 = probability of false alarm

β_0 = probability of missing signal

$Pr\{A|B\}$ = probability of A gvien that B is true

Consider two detection algorithms:

A1. Treat each beam area independently, choosing a threshold Th_1 such that:

$$\alpha_0 \geqslant Pr\;\{p_1 \geqslant Th_1 | \text{noise alone}\} \tag{1}$$

$$\beta_0 \geqslant Pr\;\{p_1 \leqslant Th_1 | \text{noise + signal } s_1\} \tag{2}$$

Since we are assuming that $f_1 \geqslant f_2$, the weakest source which may be detected is one for which $p_1 \geqslant Th_1$. Let s' be the strength of this weakest source, thus:

$$s_1 = f_1\, s' \tag{3}$$

A2. Linearly combine the powers observed in the two beams by means of a weighting function, w, and then apply a threshold Th_2. Assume the same α_0 and β_0 as in algorithm A1:

$$p_3 = wp_1 + (1 - w)p_2 \tag{4}$$

$$\alpha_0 \geqslant Pr\;\{p_3 \geqslant Th_2 | \text{noise alone}\} \tag{5}$$

$$\beta_0 \geqslant Pr\;\{p_3 \leqslant Th_2 | \text{noise + signal } s_3\} \tag{6}$$

Thus a source is detected if $p_3 \geqslant Th_2$. Let s'' be the strength of the weakest source that may be detected by this algorithm, thus:

$$s_3 = wf_1 s'' + (1 - w) f_2 s'' \tag{7}$$

The analysis which follows derives the ratio of the sensitivities of these two detection algorithms. Then the weight, w, is optimized so that s'/s'' is a maximum. Since the noise is Gaussian, p_1 and p_2 are also Gaussian with probability distribution functions:

$$P(p_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[- \frac{(p_i - \langle n \rangle)^2}{2\sigma^2} \right] \tag{8}$$

if no signal is present; and

$$P(p_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[- \frac{(p_i - \langle n \rangle - s_i)^2}{2\sigma^2} \right] \tag{9}$$

if signal s_i is present (where, $i = 1,2$). Similarly,

$$P(p_3) = \frac{1}{\sqrt{2\pi}\,\sigma_3} \exp\left[- \frac{(p_3 - \langle n \rangle)^2}{2\sigma_3^2} \right] \tag{10}$$

if no signal is present; and if signal s_3 is present,

$$P(p_3) = \frac{1}{\sqrt{2\pi}\,\sigma_3} \exp\left[-\frac{(p_3 - \langle n\rangle - s_3)^2}{2\sigma_3^2}\right] \qquad (11)$$

where we have defined:

$$\sigma_3^2 = [w^2 + (1-w)^2]\,\sigma^2 \qquad (12)$$

We are now in a position to derive an expression for the ratio of the limiting sensitivities of the two detection algorithms, s'/s'', which can then be maximized to find the optimal weight. The notation for the derivation is simplified if we define $\Phi_o(z)$:

$$\Phi_o(z) \equiv \frac{1}{\sqrt{2\pi}} \int_z^\infty \exp - \frac{t^2}{2}\, dt \qquad \text{for } z \geqslant 0$$

Assuming that $f_1 > f_2$, for strategy A1 we see from Eqs. (1) and (8) (the noise only case):

$$\alpha_0 = Pr\,\{p_1 \geqslant Th_1 \,|\, s_1 = 0\} = \Phi_o\left(\frac{Th_1 - \langle n\rangle}{\sigma}\right) \qquad (13)$$

and from Eqs. (2) and (9) (the signal present case):

$$\beta_0 = Pr\,\{p_1 \leqslant Th_1 \,|\, s_1 > 0\} = \Phi_o\left(\frac{s_1 + \langle n\rangle - Th_1}{\sigma}\right) \qquad (14)$$

On the other hand, for strategy A2 we see from Eqs. (5) and (10) (the noise only case):

$$\alpha_0 = Pr\,\{p_3 \geqslant Th_1 \,|\, s_3 = 0\} = \Phi_o\left(\frac{Th_2 - \langle n\rangle}{\sigma_3}\right) \qquad (15)$$

and from Eqs. (6) and (11) (the signal present case):

$$\beta_0 = Pr\,\{p_3 \leqslant Th_2 \,|\, s_3 > 0\} = \Phi_o\left(\frac{s_3 + \langle n\rangle - Th_2}{\sigma_3}\right) \qquad (16)$$

Now recall that we required α_0 to be the same for both algorithms. Thus Eqs. (13) and (15) are equal:

$$\left(\frac{Th_1 - \langle n\rangle}{\sigma}\right) = \left(\frac{Th_2 - \langle n\rangle}{\sigma_3}\right) \qquad (17)$$

Similarly, we require β_0 to be the same for both algorithms. Thus Eqs. (14) and (16) are also equal:

$$\left(\frac{s_1 + \langle n\rangle - Th_1}{\sigma}\right) = \left(\frac{s_3 + \langle n\rangle - Th_2}{\sigma_3}\right) \qquad (18)$$

Combining Eqs. (17) and (18), we have $s_1/s_3 = \sigma/\sigma_3$. We may use this result and Eqs. (3), (7), and (12) to find the ratio of the sensitivities of the two detection algorithms:

$$\frac{s'}{s''} = F(w) = \frac{w + \frac{f_2}{f_1}(1-w)}{\sqrt{w^2 + (1-w)^2}} \qquad (19)$$

We derive the optimal weighting strategy for combining the two responses so that the ratio of the sensitivities for the two detection algorithms is maximized by solving $dF(w)/dw = 0$:

$$w_{\text{opt}} = \frac{f_1}{(f_1 + f_2)} \qquad (20)$$

Since we have no foreknowledge of the position of the ETI source, we must employ our assumption that the signal strength is time invariant and rewrite Eq. (20) in terms of the received power and instantaneous noise power:

$$w_{\text{opt}} = \frac{(p_1 - n_1)}{(p_1 - n_1) + (p_2 - n_2)} \qquad (21)$$

The noise powers contained within the observed powers are not available to the observer and can only be estimated by calculating the expectation value. A subsequent paper (Lokshin, in preparation) will show that the optimal weighting strategy cannot be used to increase the signal to noise ratio due to the presence of cross products between the error in the estimation of the weights and the difference between the two received powers. The optimal weighting calculation does, however, give us a theoretical limit which we may use to judge practical weighting strategies.

III. Some General Results

We shall compare the theoretical limit to the enhancement of sensitivity and minimization of scalloping realized by the optimal weighting strategy to the sensitivity and scalloping (1) achieved by utilizing a single observation and (2) achieved by combining two observations with equal weights.

Suppose the sky survey could detect an ETI source of strength, s_0, if the center of the beam passes directly over the source position. What would be the sensitivity if two adjacent scans bracket the source position? Obviously, if information from only one beam area is used (strategy A1) the detectable signal is a function of the distance the source is offset from the nearer beam center (beam number 1 is assumed to be nearer):

$$s'(x_s) = s_0/f(x_s) = s_0/f_1 \tag{22}$$

Substituting Eq. (22) into (19) we find the expression for the detectable signal if one or two beam areas are used with arbitrary weighting:

$$s''(x_s) = \frac{s_0}{f(x_s) \cdot F(w)} = \frac{s_0 \sqrt{w^2 + (1-w)^2}}{wf_1 + (1-w)f_2} \tag{23}$$

This expression can now be employed to compare the peak sensitivity and scalloping as a function of x_0 for arbitrary weighting strategies, wherein it is assumed that the sky survey is carried out by scanning along parallel tracks which are separated by x_0.

Note that if $w = 1$, Eq. (23) reverts to Eq. (22), the single beam area case. This choice leaves the survey exposed to the full effect of scalloping inherent in the interscan separation. On the other hand, if we combine information from two beam areas in adjacent scans, the variation in detectable signal strength with x_s can be decreased. In the case in which the two beam areas are weighted equally without giving attention to the strength of the signal in each we find:

$$s''(x_s) = \frac{s_0 \sqrt{2}}{f_1 + f_2} \tag{24}$$

This choice of weighting function minimizes the scalloping in sensitivity at the price of degrading the peak sensitivity. Of course, application of the optimal weighting strategy achieves

$$s''(x_s) = \frac{s_0}{\sqrt{f_1^2 + f_2^2}} \tag{25}$$

Given the symmetrical beam shape, $f(x)$, these expressions may be evaluated for all x_s to determine the relative sensitivities of the three strategies and their scalloping as a function of x_0. In fact, we have done so for the case in which the beam shape is a Gaussian of arbitrary half power beam width (HPBW):

$$f(x) = \exp\left[-4 \ln(2) \left(\frac{x}{\text{HPBW}}\right)^2\right] \tag{26}$$

Figure 3 shows the relative sensitivities as a function of x_s of the three strategies for a scan separation, $x_0 = 1$ HPBW. The x-axis is shown only over the range $0 \leqslant x_s \leqslant x_0$ since the response is symmetrical about either limit. For the case in which data from only a single beam area is used, scalloping is 3 dB. For the case in which data from two neighboring beam areas are combined with equal weight, scalloping is reduced to 0.34 dB. This is achieved at the cost of a loss in peak sensitivity of 1.17 dB relative to the single beam result due to averaging data containing almost no signal with data having the maximum signal to noise ratio. For the case in which the optimal weight has been used, scalloping is reduced to 1.5 dB while peak sensitivity is enhanced 0.01 dB relative to the single beam result.

Figure 4 shows the relative sensitivities as a function of x_s for a scan separation, $x_0 = 0.75$ HPBW. Scalloping for the single beam case is reduced to 1.7 dB whereas scalloping for the equal weight case is increased to 0.48 dB and the loss in peak sensitivity relative to the single beam result is decreased to 0.2 dB. Note that the larger scalloping is due to a greater enhancement of sensitivity between the beam areas than at their centers. This effect was just barely noticable in the preceeding figure. Scalloping for the optimal weight case is reduced to 0.3 dB and the peak sensitivity is enhanced by 0.1 dB relative to the single beam result.

Figure 5 shows the variation of scalloping as a function of x_0 for the three weighting strategies. For large x_0 the improvement in scalloping over that resulting from a single beam is 3 dB for equal weighting and 1.5 dB for optimal weighting. Note, however, that minima occur in the scalloping for the equal weighting and optimal weighting cases. For a scan separation smaller than 0.95 HPBW, the sensitivity to a source located between equally weighted scans increases faster than the sensitivity to a source located at the center of either scan as the scan separation is decreased. Thus the nonuniformity in sensitivity increases with decreasing scan separation until a substantial overlap is achieved. A similar phenomenon occurs for the optimally weighted scans after x_0 shrinks below 0.65 HPBW.

Figure 6 shows the variation of the peak sensitivity as a function of x_0 relative to that achieved using only a single beam. The optimal weighting strategy always achieves a better sensitivity than a single beam, albeit the improvement is not very great for $x_0 \geqslant 0.65$ HPBW. The equal weighting strategy peak sensitivity has already dropped halfway to its asymptotic minimum at $x_0 \approx 0.85$ HPBW.

The variations in scalloping and peak sensitivity shown in Figs. 5 and 6 are key considerations in the design of a sky survey strategy and will impact the survey sensitivity, given the constraints on antenna time and available memory and processing power. The sensitivity of the survey can be increased by (1) dwelling longer in each beam area, (2) decreasing the scan separation, (3) combining data from neighboring beam areas, or (4) combining all three options. Each option has its price, however.

In the first option, the time to complete the survey increases as the square of the ratio of new to old sensitivities, but the scalloping is not affected. In the second option, the time increases directly as the ratio of old to new scan separations increases and the scalloping is reduced. In the third option, survey time is not affected and scalloping is reduced, but memory and processing requirements increase and peak sensitivity may suffer depending upon the manner of combination chosen. The fourth option will always be chosen, but the mix will vary depending upon the constraints under which the survey will operate.

IV. The Effect of Scan Separation in the Case of Uniformly Distributed Transmitters and Fixed Survey Time

Plausible arguments may be advanced for any number of assumptions about the spatial, power, duty cycle, and transmitted frequency distribution functions for signals of ETI origin. Given a set of assumptions, a survey may be tailored to maximize the probability that it will detect a signal of that class. Many reasonable scenarios have been advanced in the literature, but the great advantage of an all sky survey lies in the fact that it incorporates the fewest *a priori* assumptions.

We now consider the impact of scan separation and weighting strategy upon the probability of detecting an ETI signal. The calculation requires that some assumptions be made about the distribution of sources and the manner in which the survey will be carried out. In the light of real life constraints for a sky survey and our state of ignorance concerning possible ETI sources, we shall follow in the footsteps of Drake (Ref. 2) and Gulkis (Ref. 3) and assume that:

(a) An $M \times N$ HPBW^2 area of the sky is to be surveyed in a fixed time, T.

(b) The transmitters are of equal strength and are distributed uniformly throughout space.

Suppose that our hypothetical sources each have an effective isotropic radiation power, P, and the boresight gain of the antenna is G. If a particular source is a distance, R, from earth and is displaced relative to the boresight by a distance, x, the flux seen by the receiving system in one beam area which is due to the source is

$$s(x) = \frac{G \cdot P}{4\pi R^2} \cdot f(x) \tag{27}$$

If the source is to be detected, this flux must be greater than or equal to the minimum detectable flux, s'. Thus, the source will be detected if it is closer than the distance R_m:

$$R_m(x) = \left[\frac{G \cdot P}{4\pi s'} \cdot f(x)\right]^{1/2} \tag{28}$$

Equation (19) may be substituted into (28) to find the expression for the maximum distance if the weighted data from two beam areas are used. We must keep in mind that $f(x)$ is really f_1 in Eq. (19), and we must explicitly show the dependence upon the integration time, τ, since $s_0 \sim \tau^{-1/2}$:

$$R_m(x) \sim \left[\frac{G \cdot P}{s''} \cdot f(x) \cdot F(w) \cdot \sqrt{\tau}\right]^{1/2} \tag{29}$$

The number of detectable transmitters (and thus the probability of detecting one) is proportional to the volume of space observed:

$$\delta Pr\{\text{detection}\} \sim \delta V = R_m^3 \, \delta\Omega \tag{30}$$

Substituting Eq. (29) into (30), we have:

$$\delta Pr\{\text{detection}\} \sim \left[\frac{GP}{s''}\right]^{3/2} \tau^{3/4} \, \{f(x) \cdot F(w)\}^{3/2} \, \delta\Omega \tag{31}$$

Assumption (a) may now be applied as a constraint so that the effect of the scan separation, x_0, upon the probability of detecting a signal may be evaluated. The simplest survey strategy entails making scans which are N HPBWs long, stepping by x_0 in the orthogonal coordinate until a distance of M HPBWs is covered. Thus the number of scans is equal to M/x_0. The integration time is set equal to some fraction of the amount of time required to scan through one HPBW, and we shall assume here that the fraction is unity. Thus the total time allowed for the survey (assuming zero time between scans) must be:

$$T = \frac{M}{x_0} \cdot N \cdot \tau \tag{32}$$

Solving for τ, we may now integrate Eq. (31) over all x_s to derive an expression for the probability of detecting a signal as a function of scan separation:

$$Pr\{\det | x_0\} \sim \left[\frac{MN}{x_0}\right]^{1/4} \left[\frac{GP\sqrt{T}}{s''}\right]^{3/2} \times 2\int_0^{1/2x_0} \left[\frac{wf(x) + (1-w)f(x_0 - x)}{\sqrt{w^2 + (1-w)^2}}\right]^{3/2} dx \tag{33}$$

Figure 7 shows the variation of a normalized $Pr\{\text{detection} | x_0\}$ with x_0 for the three weighting strategies. The normalization is chosen so that the maximum relative probability for the optimal weighting strategy is unity. As x_0 becomes larger, the scans can be slowed down and still allow the survey to cover the same area of sky in the given time limit. However, the scalloping in sensitivity increases with x_0 and degrades the probability of detection. For very small values of x_0, the optimal weighting and equal weighting strategy improve the probability over that achieved by the single beam strategy by the expected ratio of $2^{3/4}$ due to the effective doubling of integration time on source. As the scan separation increases the single beam area strategy approaches the optimal result and the equal weighting strategy falls off to the expected ratio of 1/2 due to doubling of the noise.

The relative probabilities of detection for the single beam strategy and the equal weighting strategy are equal for a scan separation of about 0.8 HPBW, and are degraded relative to the optimal result by approximately 15%. The probability of detection achieved by the equal weighting strategy peaks at a scan separation of 0.6 HPBW, and it is degraded from the optimal peak by 8%. The probability of detection achieved by using a single beam area peaks at a scan separation of 1.3 HPBW, and it is degraded from the optimal peak by 5%. It is clear that the scalloping allowable in a survey will depend upon the assumptions of the designers of the search strategy.

V. Suggestions for Further Analysis

It is possible to extend the foregoing analysis to cover the general case of combining N beam areas. Of more immediate concern, however, is an extension to the case of a non-Gaussian noise statistic. The contemplated high speed all sky survey will operate in this domain due to the small number ($4 < n < 100$) of independent samples which will be combined before thresholding. A series of papers is in preparation which will cover this topic.

Acknowledgments

The authors would like to thank S. Gulkis and M. Klein for reading the manuscript and making helpful suggestions for its improvement.

References

1. Gulkis, S., Olsen, E. T., and Tarter, J., "A Bimodal Search Strategy for SETI," *Strategies for the Search for Life in the Universe*, edited by M. D. Papagiannis. The Proceedings of the IAU General Assembly, Vol. 83, pp. 93-105. D. Reidel, Dodrecht, Holland, the Netherlands, 1980.
2. Drake, F. D., "Estimates of the Relative Probability of Success of the SETI Search Program," *SETI Science Working Group Report*, edited by F. Drake, J. H. Wolfe, C. L. Seeger. NASA Technical Report, p. 67, October 1983.
3. Gulkis, S. "Note on the Optimal Search Strategy for Uniformly Distributed CW Transmitters," *TDA Progress Report 42-77*, January-March 1984, Jet Propulsion Laboratory, Pasadena, California.

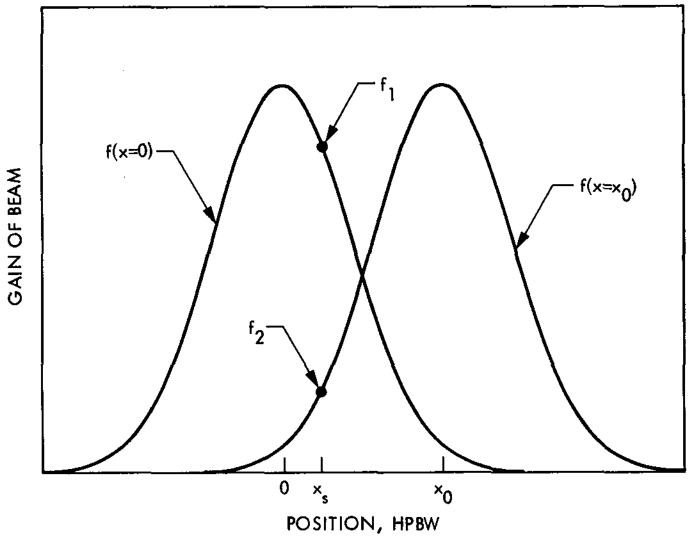


Fig. 1. One dimensional model of survey geometry

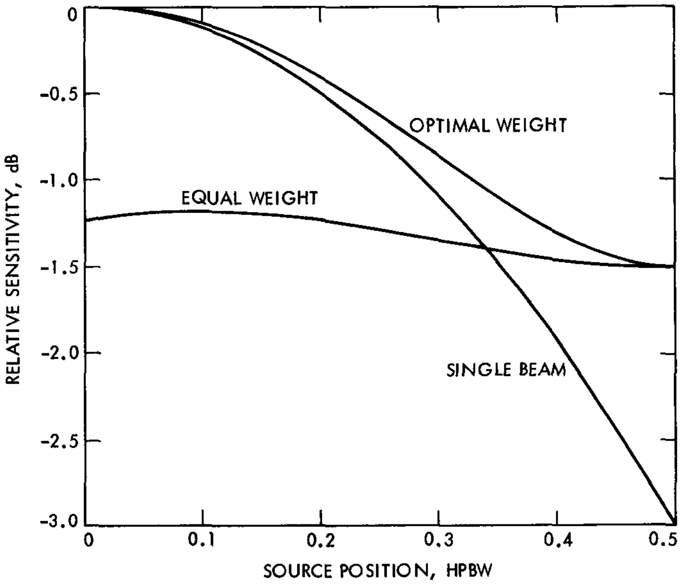


Fig. 3. Relative sensitivity as a function of source position for a scan separation of 1.0 HPBW

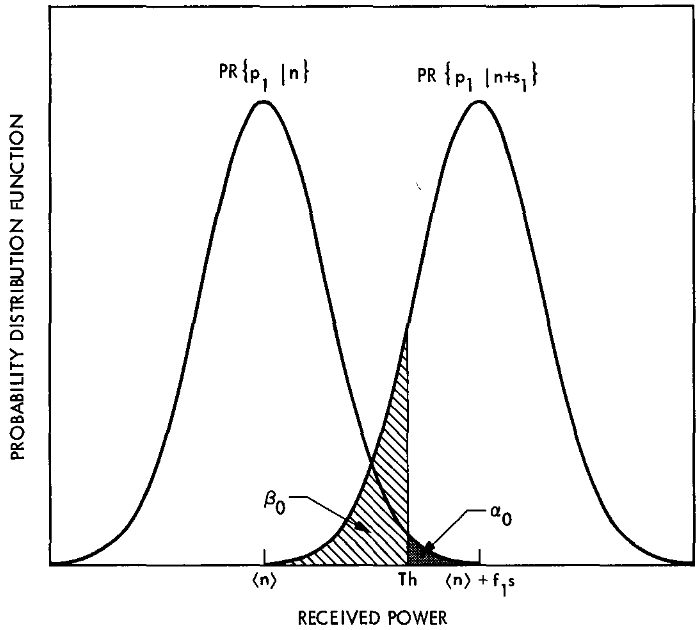


Fig. 2. Probability distribution functions of received power due to noise alone and due to noise and source

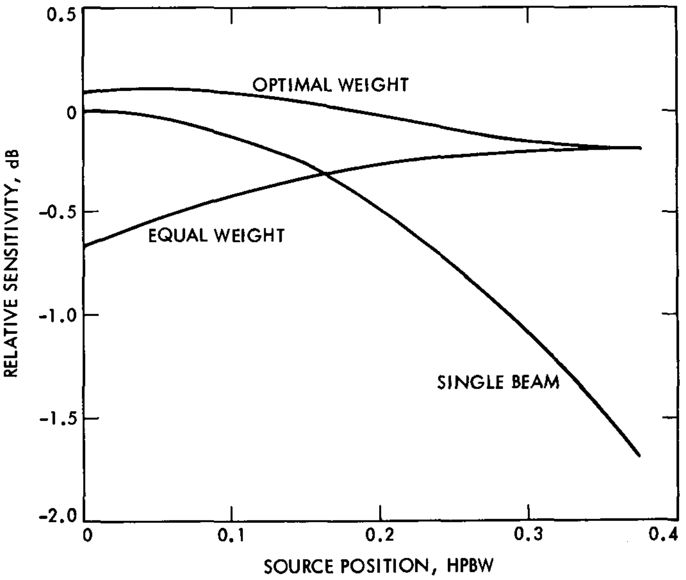


Fig. 4. Relative sensitivity as a function of source position for a scan separation of 0.75 HPBW

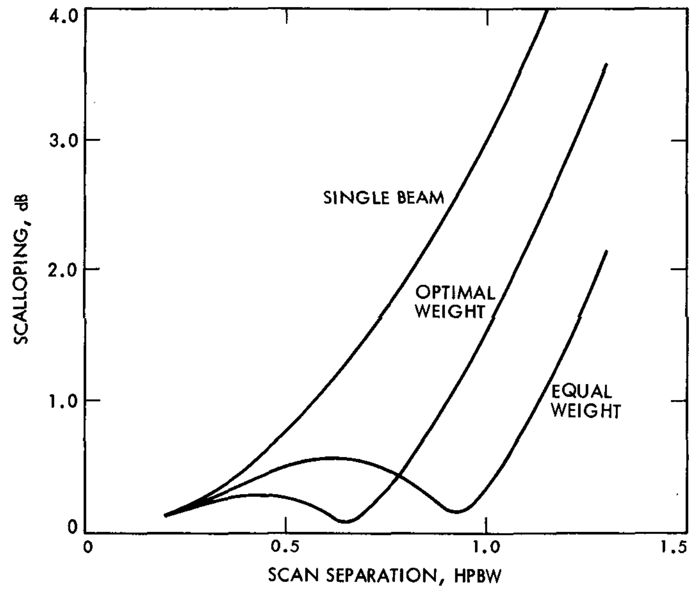


Fig. 5. Scalloping as a function of scan separation

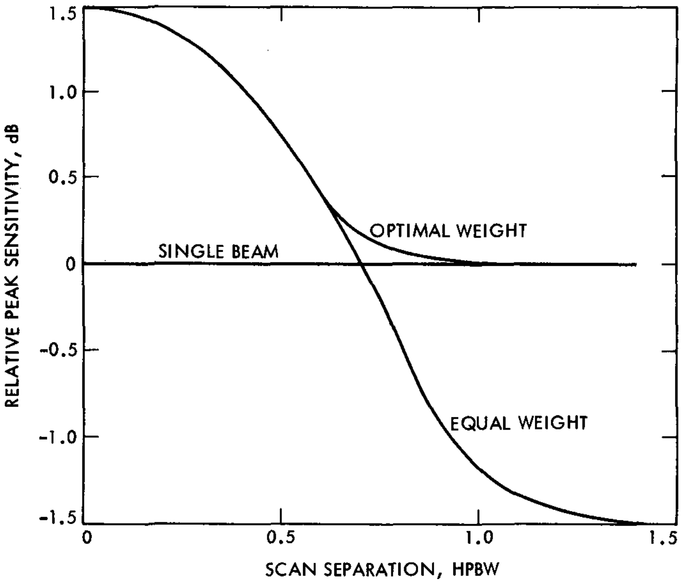


Fig. 6. Relative peak sensitivity as a function of scan separation

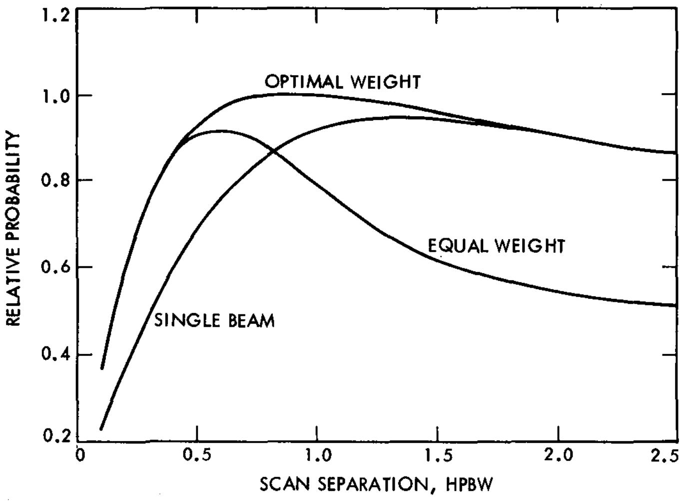


Fig. 7. Normalized relative probability to detect ETI as a function of scan separation