This document has been converted from the original publication:
Barry, K. M., Cavers, D. A. and Kneale, C. W., 1975, Report on recommended standards for digital tape formats: Geophysics, 40, no. 02, 344-352.

RECOMMENDED STANDARDS FOR DIGITAL TAPE FORMATS[1]

K. M. BARRY[2], D. A. CAVERS[3], AND C. W. KNEALE[4]

INTRODUCTION

Recently, a new demand for demultiplexed formats has arisen in the seismic industry due to the utilization of minicomputers in digital field recording systems, and because of a growing need to standardize an acceptable data exchange format.

In 1973 a subcommittee of the SEG committee on Technical Standards was organized to gather information and develop a nine-track, ½-inch tape, demultiplexed format for industry acceptance. Guidelines set for this new format were based on prior work and on the SEG Exchange Tape Format (Northwood et. al, 1967). As a result of the subcommittee's effort based on suggestions from industry personnel, the following demultiplexed format recommendations are made.

The present SEG Exchange Tape Format is often referred to as the SEG "Ex" Format. Because of this, it is recommended that the new demultiplexed format be designated the "SEG Y Format." The Technical Standards committee has elected to withdraw support of the SEG "Ex" Format.

The SEG Y Format was developed for application to computer field equipment and in the present data processing center with flexibility for expansion as new ideas are introduced. Current information for standardization is placed in the "fixed" portion of the format, while new ideas can be added to the unassigned portions later as expansion becomes necessary.

It is assumed that this format will accommodate the majority of field and office procedures and the techniques presently utilized.

FORMAT SPECIFICATION

The following general information describes the recommended demultiplexed format (Figure 1):

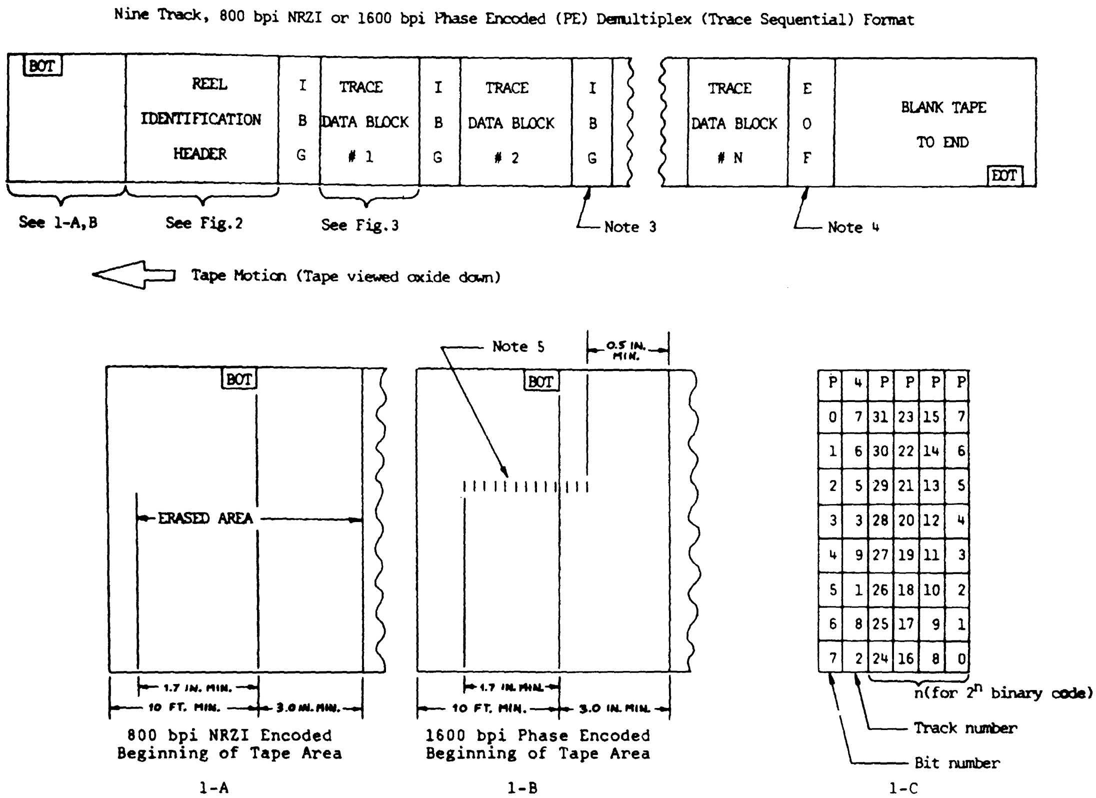
1) Tape specifications, track dimensions and numbering, and all other applicable specifications shall be in accordance with IBM Form GA 22-6862 entitled "IBM 2400-Series Magnetic Tape Units Original Equipment Manufacturers' Information".

At the present time, IBM has proposed an American National Standard for the 6250 CPI group coded recording format. Should this format be used within the geophysical industry, the applicable IBM specifications would apply. The additional formatting required by this proposed method is a function of the hardware and thus becomes transparent to the user.

2) Either the NRZI encoded data at 800-bpi density, or the phase encoded (PE) data at 1600-bpi density may be used for recording.
3) All data values are written in two's complement except the 320bit floating point format, Figure 3-A, which is sign, characteristic, and fractional part.
4) Data values are written in eight-bit bytes with vertical parity odd.

---

[1] ©1975 Society of Exploration Geophysicists. All rights reserved.
This report is the work of the Subcommittee on Demultiplexed Field Tape Formats of the SEG Technical Standards Committee. Manuscript received by the Editor October 7, 1974.
[2] Subcommittee Chairman, Teledyne Exploration, Houston. Tex. 77036.
[3] Subcommittee Member, Gulf Research & Development Co., Houston, Tex. 77036.
[4] Subcommittee Member, Texas Instruments Inc., Houston, Tex. 77001.



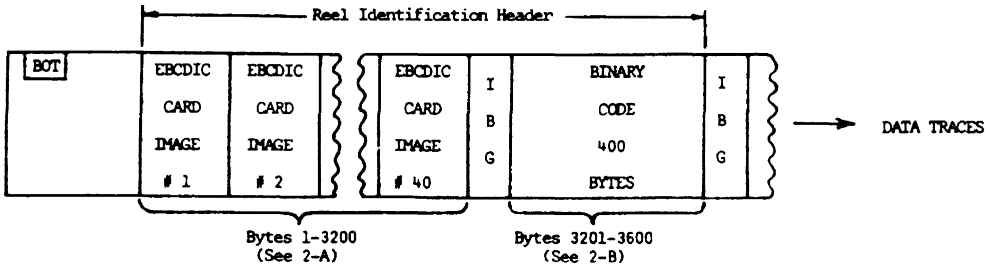
P	4	P	P	P	P
0	7	31	23	15	7
1	6	30	22	14	6
2	5	29	21	13	5
3	3	28	20	12	4
4	9	27	19	11	3
5	1	26	18	10	2
6	8	25	17	9	1
7	2	24	16	8	0

Fig. 1. Recommended demultiplexed format.

Notes:

1. **Preamble-Proceeds each of the 45 blocks within the reel identification header and. each trace data block when 1600 bpi PE is used. Consists of 40 all-zero bytes followed by one all-ones byte.**
2. **Postamble-Follows each of the 45 blocks within the reel identification header and each trace data block when 1600 bpi PE is used. Consists of one all-ones byte followed by 40 all-zero bytes.**
3. **Interblock Gap (IBG)-Consists of 0.6" nominal, 0,5" minimum.**
4. **End of file (EQF)-Consists of an IBG followed by:**
   a) **PE tape mark having 80 flux reversals at 3200 fci in bit numbers F,0,2,5,6, and 7. Bits 1,3, and 4 are dc-erased, or**
   b) **NRZI tape mark having two bytes with one bits in bit numbers 3,6, and 7 separated by seven all-zero bytes**
5. **PE Identification Burst-Consists of 1600 flux reversals per inch in bit number P; all other tracks are erased.**

5) Definitions:
   a) *Interblock gap (IBG)* – Consists of erased tape for a distance of 0.6 inches nominal, 0.5 inches minimum.
   b) *End of file (EOF)* – Consists of the 800-bpi NRZI tape mark or the 1600-bpi tape mark character, as appropriate, preceded by a standard IBG.
   c) *Erased tape* – The tape is magnetized, full width, in a direction such that the rim end of the tape is a north-seeking pole. The read-back signal from such an area shall be less than 4 percent of the average signal level at 3200 flux reversals per inch.
   d) *PE identification burst* – Consists of 1600 flux reversals per inch in bit number P with all other traces DC erased. This burst is written beginning at least 1.7 inches before the trailing edge of the beginning of tape (BOT) reflective marker and continuing past the trailing edge of the marker, but ending at least 0.5 inches before the first block.
   e) *Block* – Continuous recorded information, preceded and followed by a standard IBG. In PE (1600 bpi), a preamble precedes each block and a postamble follows each block.
   f) *Preamble* – Consists of 41 bytes, 40 of which contain zero bits in all tracks; these



2-A EBCDIC CARD IMAGES Free form coding, left justified – 40 card images, 80 bytes per card, card image numbers 23-39 unassigned, for optional information.

are followed by a single byte containing one bits in all tracks.

g) *Postamble* – Consists of 41 bytes of which the first byte contains one bits in all tracks; it is followed by 40 bytes containing zero bits in all tracks.

h) *Two's complement* – Positive values are the true binary number. Negative values are obtained by inverting each bit of the positive binary number and adding one (1) to the least significant bit position.

6) The seismic reel is divided into the reel identification header and the trace data blocks. The reel identification header section contains identification information pertaining to the entire reel and is subdivided into two blocks, the first containing 3200 bytes of EBCDIC card image information (equivalent of 40 cards) and the second consisting of 400 bytes of binary information. These two blocks of the reel identification header are separated from each other by an IBG. Each trace data block contains a trace identification header and the data values of the seismic channel or auxiliary channels. The reel identification header and the first trace data block are separated by an IBG.

7) Each seismic-trace data block is ungapped and is written in demultiplexed format with each trace data block being separated from the next by an IBG. The last trace data block on the reel is followed by one (or more) EOF>

8) When recorded 800 bpi (NRZI), the first block of the reel identification header begins at least 3.0 inches past the trailing edge of the BOT marker.
9) The following conventions pertain to the reel and trace identification headers:
   a) All binary entries are right justified. All EBCDIC entries are left justified.
   b) All times are in milliseconds with the exception of the sample interval which is designated in microseconds.
   c) All frequencies are in hertz.
   d) All frequency slopes are in dB/octave.
   e) All distances (lengths) are in feet or meters, and these systems are not mixed within a reel. The distance or measurement system used is specified in card image 7 and in bytes 3255-3256 of the reel identification header.
   f) A scaler may be applied to certain distance measurements where greater precision is required. See bytes 69-70 and 71-72 of the trace identification header.
   g) The energy source and geophone group coordinates designated in bytes 73-88 of the

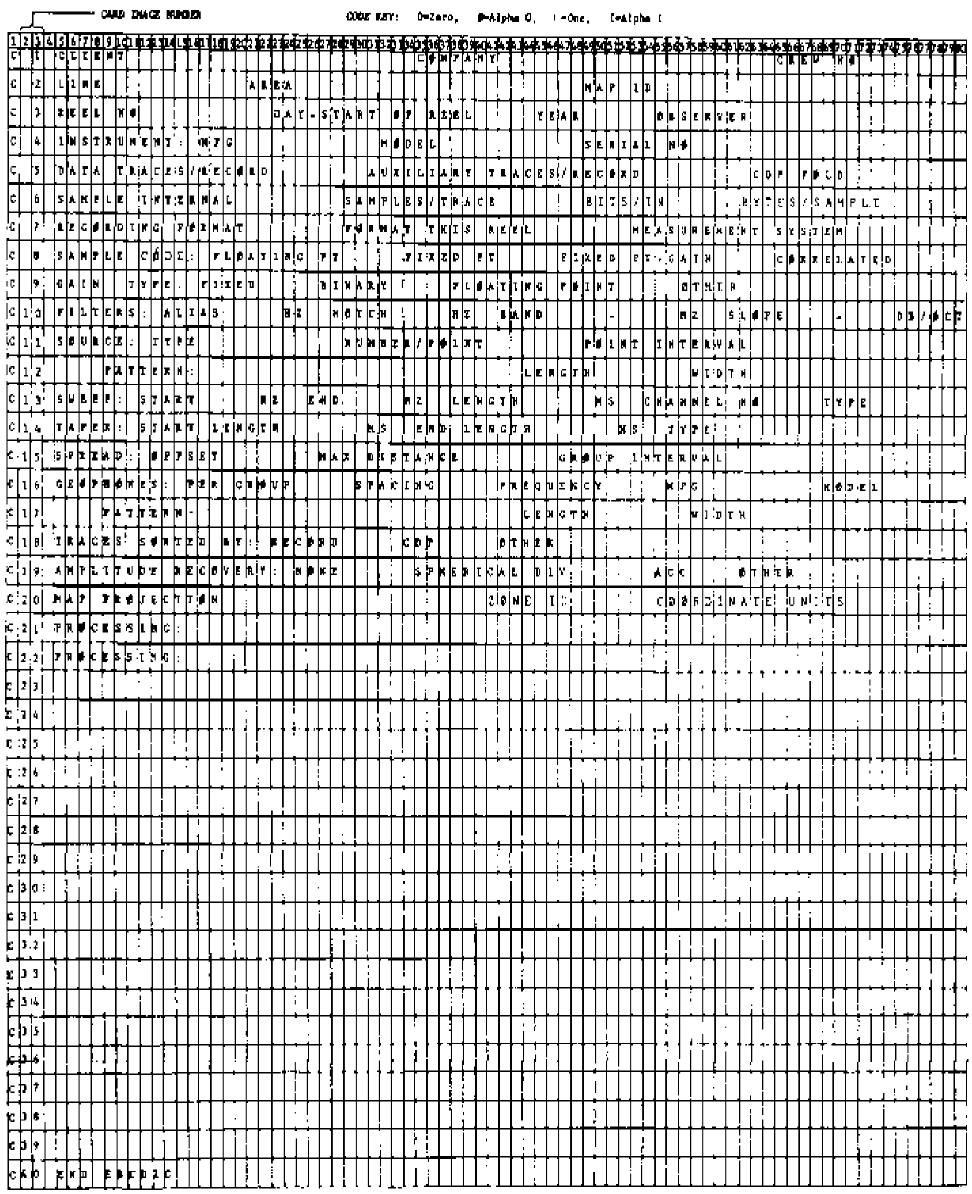


Fig. 2A. Reel identification header. Part 1, the EBCDIC card image block.

trace identification header can be measured in either length or latitude and longitude. The measurement unit used is specified in bytes 89-90 of the trace header. For the latitude/longitude system, the coordinate values are expressed in seconds of arc.

h) All velocities are in feet per second or meters per second, and these units are not mixed within a reel.

i) Elevation is represented by "+" above "—" below mean sea level.

10) The binary coded information convention is defined in Figure 1-C.

DESCRIPTION OF REEL IDENTIFICATION HEADER

The reel identification header (Figure 2) consists of 3600 bytes and is divided into two parts:

1) The card image EBCDIC block (3200) bytes— 40 cards equivalent) followed by an IBG.
2) The binary coded block (400 bytes) followed by an IBG.

The EBCDIC part of the reel header describes the data from a line of shotpoints in a fixed specified format consisting of 40 card images with each image containing 80 bytes. All unused card image characters are EBCDIC Blank. Card image numbers 23 through 39 are unassigned for optional use. Each card image should contain the character "C" in the first card column. Each 80 bytes would yield one line of format print to produce the form shown in Figure 2-A.

The binary coded section of the reel header consists of 400 bytes of information common to the seismic data on the related reel as shown in Figure 2-B. There are 60 bytes assigned; 340 are unassigned for optional use.

There are certain bytes of information that may not apply to a particular recording or processing procedure. It is strongly recommended that bytes designated with an asterisk (*) in Figures 2-B and 3-E always contain the required information

The data in the reel identification header could be printed and edited prior to the actual input of seismic data for processing. A complete header listing of both the EBCDIC and binary parts would accompany an exchange tape and serve as a table of contents and summary of specifications for that reel* of seismic data. No more than one line of seismic data is permitted on any one reel. Additional reels would be used for long lines, and each reel must start with a reel identification header.

DESCRIPTION OF THE TRACE DATA BLOCK

Each trace data block (Figure 3) consists of a fixed 240-byte trace identification header and the seismic trace data block is separated from the next by an IBG. The trace header is written in binary code (refer to Figure 1-C for the binary code information) and is detailed in Figure 3-E.

The trace data samples can be written in one of the four data sample formats described in Figures 3-A, 3-B, 3-C, and 3-D. The trace data format for each reel is identified in bytes 3225-3226 of the reel identification header. Only one data sample format is permitted within each reel.

Figure 3-A details a 32-bit, floating point format in which each data value of a seismic channel is recorded in four successive bytes, in IBM compatible floating point notation as defined in IBM Form GA 22-6821.

The four bytes form a 32-bit word consisting of the sign bit Q_S, a seven-bit characteristic Q_C, and a 24-bit fraction Q_F. Q_S indicates signal polarity and is a one for a negative value. Q_C signifies a power of 16 expressed in excess 64 binary notation allowing both negative and positive powers of 16 to be represented by a true number. Q_F is a six hexadecimal digit (24 amplitude recovery can be described in the binary bit) number with a radix point to the left of the significant digit. The data value represented by a floating point number is

$$\mathbf{Q_S \; X \; 16^{(O\,C-64)} \; X \; Q_F}$$

Figure 3-B details a 32-bit, fixed point format and each data value of a seismic channel is recorded in four successive bytes. This format consists of a sign bit $\mathbf{Q_S}$ (one represents negative) and 31 data bits $\mathbf{Q_D}$ with a radix point at the right of the least significant digit.

Figure 3-C represents a 16-bit, fixed point format, and each data value of a seismic channel is recorded in two successive bytes. This format is similar to figure 3-B except there are 15 data bits $\mathbf{Q_D}$.

Figure 3-D represents a 32-bit, fixed point format with gain values. The first byte of this format is all zeros. The second byte provides eight available gain bits 2^0 through 2^7. The last two bytes are identical to Figure 3-C.

In all four data formats, the channel or trace data should represent the absolute input voltage at the recording instrument. The 32-bit, floating point field format defined as the SEG C (Meiners et al, 1972) comprehends the input voltage level. The fixed point formats 3-B and 3-C require a trace weighting factor (trace identification header, bytes 169-170), defined as 2^{-n} volts for the least significant bit, to comprehend the absolute input voltage level.

In cases where processing parameters such as amplitude recovery are present, the type of amplitude recovery can be described in the appropriate reel identification header sections, and the algorithm described in the unassigned portions.

CONCLUSION

Individual oil companies and contractors may be convinced of their own format's merits, but the use of this recommended exchange demultiplexed format must be given serious consideration in order to achieve some level of industry standardization. Such thought and many suggestions from users have been utilized in establishing a flexible format that yields specifics and can be used by all companies in the industry.

Adoption and use of this format will save substantial sums of money in computer time and programming effort in the future.

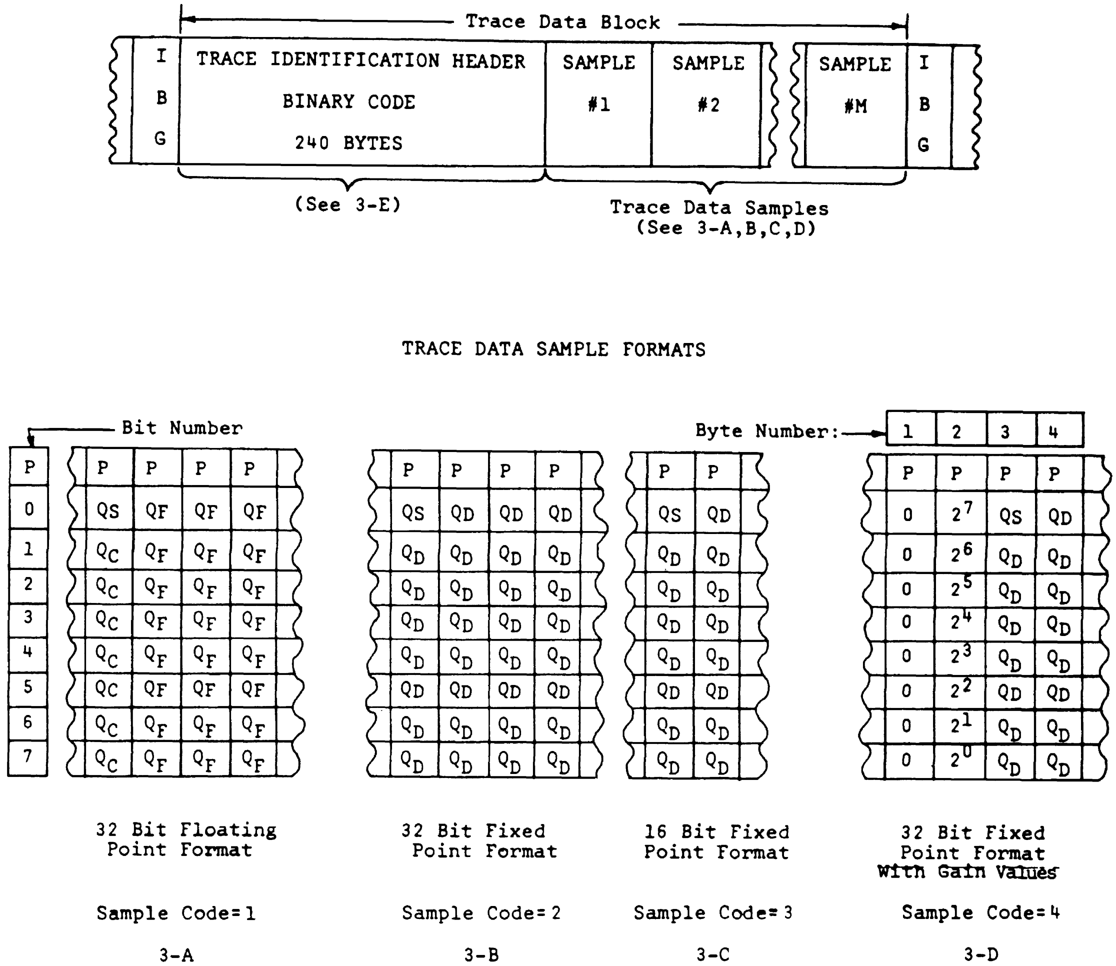
ACKNOWLEDGEMENTS

Grateful appreciation goes to many companies and individuals for their suggestions at the start of the subcommittees' work and for their final recommendations. We are also for the assistance of Fred Tischler, Texas Instruments, who was the original subcommittee chairman.

REFERENCES

Meiners E. P., Lenz, L. L., Dalby, A. E., and Hornsby, J. M., 1972, Recommended standards for digital tape formats: Geophysics, v. 37, p. 36-44.

Northwood E. J., Wisinger, R. C., and Bradley J. J., 1967, Recommended standards for digital tape formats: Geophysics, v. 32, p. 1073-1084.



32 Bit Floating Point Format	32 8it Fixed Point Format	16 Bit Fixed Point Format	32 Bit Fixed Point Format with Gain Values
Sample Code=1	Sample Code=2	Sample Code=3	Sample Code=4
3-A	3-B	3-C	3-D

<u>NOTE</u>: Least significant bit is always in bit position 7 of byte 4 (or byte 2 for 3-C).

Q_S = Sign bit
Q_C = Characteristic
Q_F = Fraction
Q_D = Data bits

FIG. 3A-D. Trace data block. Four data sample options.

2-B. BINARY CODE-Right justified

Byte Numbers		Description		
3201-3204		Job identification number.		
3205-3208	*	Line number (only one line per reel).		
3209-3212	*	Reel number.		
3213-3214	*	Number of data traces per record (includes dummy and zero traces inserted to fill out the record or common depth point).		
3215-3216	*	Number of auxiliary traces per record (includes sweep, timing, gain, sync, and all other nondata traces).		
3217-3218	*	Sample interval in µsec (for this reel of data).	Designated in microseconds to accommodate sample intervals less	
3219-3220		Sample interval in µsec (for original field recording).	than one millisecond.	
3221-3222	*	Number of samples per data trace (for this reel of data).		
3223-3224		Number of samples per data trace (for original field recording).		
3225-3226	*	Data sample format code:	1 = floating point (4 bytes) 2 = fixed point (4 bytes.)	3 = fixed point (2 bytes) 4 = fixed point w/gain code
		Auxiliary traces use the same number of bytes per sample. (4 bytes)		
3227-3228	*	CDP fold (expected number of data traces per CDP ensemble).		
3229-3230		Trace sorting code:	1 = as recorded (no sorting) 2 = CDP ensemble	3 = single fold continuous profile 4 = horizontally stacked
3231-3232		Vertical sum code:	1 = no sum, 2 = two sum, ..., N = N sum (N = 32,767)	
3233-3234		Sweep frequency at start.		
3235-3236		Sweep frequency at end.		
3237-3238		Sweep length (msec).		
3239-3240		Sweep type code:	1= linear 2= parabolic	3 = exponential 4 = other
3241-3242		Trace number of sweep channel.		
3243-3244		Sweep trace taper length in msec at start if tapered (the taper starts at zero time and is effective for this length).		
3245-3246		Sweep trace taper length in msec at end (the ending taper starts at sweep length minus the taper length at end).		
3247-3248		Taper type:	1 = linear 2 = cos2	3 = other
3249-3250		Correlated data traces:	1 = no	2 = yes
3251-3252		Binary gain recovered:	1 = yes	2 = no
3253-3254		Amplitude recovery method:	1 = none 2 = spherical divergence	3 = AGC 4 = other
3255-3256		Measurement system:	1 = meters	2 = feet
3257-3258		Impulse signal	1 = Increase in pressure or upward geophone case movement gives negative number on tape.	
		Polarity	2 = Increase in pressure or upward geophone case movement gives positive number on tape.	
3259-3260		Vibratory polarity code:	Seismic signal lags pilot signal by:	
		1 =	337.5° to 22.5°	
		2 =	22.5° to 67.5°	
		3 =	67.5° to 112.5°	
		4 =	112.5° to 157.5°	
		5 =	157.5° to 202.5°	
		6 =	202.5° to 247.5°	
		7 =	247.5° to 292.5°	
		8 =	292.5° to 337.5°	
3261-3600		Unassigned – for optional information.		

*Strongly recommended that this information always be recorded.

Byte Numbers	Description
1 - 4	* Trace sequence number within line--numbers continue to increase if additional reels are required on same line.
5 - 8	Trace sequence number within reel--each reel starts with trace number one.
9 –12	* Original field record number.
13-16	* Trace number within the original field record.
17-20	Energy source point number--used when more than one record occurs at the same effective surface location.
21-24	CDP ensemble number
25-28	Trace number within the CDP ensemble--each ensemble starts with trace number one.
29-30	* Trace identification code: 1 = seismic data 4 = time break 7 = timing 2 = dead 5 = uphole 8 = water break 3 = dummy 6 = sweep 9---- N = optional use (N = 32,767)
31-32	Number of vertically summed traces yielding this trace. (1 is one trace, 2 is two summed traces, etc.)
33-34	Number of horizontally stacked traces yielding this trace. (1 is one trace, 2 is two stacked traces, etc.)
35-36	Data use: 1 = production. 2 = test.
37-40	Distance from source point to receiver group (negative if opposite to direction in which line is shot).
41-44	Receiver group elevation; all elevations above sea level are positive and below sea level are negative.
45-48	Surface elevation at source.
49-52	Source depth below surface (a positive number).
53-56	Datum elevation at receiver group.
57-60	Datum elevation at source.
61-64	Water depth at source.
65- 68	Water depth at group.
69-70	Scaler to be applied to all elevations and depths specified in bytes 41-68 to give the real value. Scaler = 1, +10, +100, +1000, or +10,000. If positive, scaler is used as a multiplier; if negative, scaler is used as a divisor.
71-72	Scaler to be applied to all coordinates specified in bytes 73-88 to give the real value. Scaler = 1, +10, +100, +1000, or +10,000. If positive, scaler is used as a multiplier: if negative, scaler is used as divisor.
73-76	Source coordinate - X.
77-80	Source coordinate - Y.
81-84	Group coordinate - X.
85-88	Group coordinate - Y.
	If the coordinate units are in seconds of arc, the X values represent longitude and the Y values latitude. A positive value designates the number of seconds east of Greenwich Meridian or north of the equator and a negative value designates the number of seconds south or west.
89-90	Coordinate units: 1 = length (meters or feet). 2 = seconds of arc.
91-92	Weathering velocity.
93-94	Subweathering velocity.
95-96	Uphole time at source.
97-98	Uphole time at group.
99-100	Source static correction.
101-102	Group static correction.
103-104	Total static applied. (Zero if no static has been applied,)

FIG. 3E. Trace identification header written in binary code.

Byte Numbers	Description
105-106	Lag time A. Time in ms. between end of 240-byte trace identification header and time break. Positive if time break occurs after end of header, negative if time break occurs before end of header. Time break is defined as the initiation pulse which may be recorded on an auxiliary trace or as otherwise specified by the recording system.
107-108	Lag Time B. Time in ms. between time break and the initiation time of the energy source. May be positive or negative.
109-110	Delay according time. Time in ms. between initiation time of energy source and time when recording of data samples begins. (for deep water work if data recording does not start at zero time.)
111-112	brute time--start.
113-114	Mute time--end.
115-116	* Number of samples in this trace.
117-118	* Sample interval in µsec for this trace.
119-120	Gain type of field instruments: 1 = fixed. 2 = binary. 3 = floating point. 4 --- N = optional use.
121-122	Instrument gain constant.
123-124	Instrument early or initial gain (dB).
125-126	Correlated: 1 = no. 2 = yes .
127-128	Sweep frequency at start.
129-130	Sweep frequency at end.
131-132	Sweep length in ms.
133-134	Sweep type: 1 = linear. 2 = parabolic. 3 = exponential. 4 = other.
135-136	Sweep trace taper length at start in ms.
137-138	Sweep trace taper length at end in ms.
139-140	Taper type: 1 = linear. 2 = cos2. 3 = other.
141-142	Alias filter frequency, if used.
143-144	Alias filter slope
145-146	Notch filter frequency, if used.
147-148	Notch filter slope.
149-150	Low cut frequency, if used .
151-152	High cut frequency, if used .
153-154	Low cut slope
155-156	High cut slope
157-158	Year data recorded .
159-160	Day of year.
161-162	Hour of day (24 hour clock)
163-164	Minute of hour.
165-166	Second of minute.
167-168	Time basis code: I = local. 2 = GMT. 3 = other.
169-170	Trace weighting factor--defined as 2-N volts for the least significant bit. (N = 0, 1, ... 32, 767.)
171-172	Geophone group number of roll switch position one.
173-174	Geophone group number of trace number one within original field record .
175-176	Geophone group number of last trace within original field record.
177-178	Gap size (total number of groups dropped).
179-180	Overtravel associated with taper at beginning or end of line: I = down (or behind). 2 = up (or ahead).
181-240	Unassigned—for optional information.

* Strongly recommended that this information always be recorded.

FIG. 3E. Trace identification header written in binary code (cont.)