AVALON SCIENCES LIMITED

MIRF7 Specification

Commercial in Confidence

This document may be used for reference but its circulation and updating is not controlled.
This may not be the latest version currently in force.

Avalon Sciences Limited
Avalon House, Somerton Business Park, Bancombe Road
Somerton, Somerset, UK, TA11 6SB

Contents

Amendment record

Issue:	Date:	Prepared:	Comments/Changes:
1	11th July 2019	P Royds	First Draft in line with TD982

1 Introduction

MIRF (Media Independent Recording Format) is designed to simplify the management of VSP and Microseismic data recorded on a PC based data acquisition system. MIRF7 is a superset of the original MIRF specification, providing for the additional requirements of the current generation of data acquisition systems and microseismic recording.

A MIRF record is a PC compatible file, so can easily be transferred between different types of media using standard PC file utilities.

2 Changes introduced at version 7

1. The High Side Indicator Roll field is now reported in the range 0-360 degrees instead of -180 to +180.
2. The High Side Indicator Inclination field, is reported in the range 0-180 degrees. This was also the case in MIRF 6, but the document incorrectly stated the range to be -90 to +90 degrees.
3. Document updated to describe the GPRMC GPS field in the header block, an undocumented feature of the MIRF6 specification.

3 Note about UTC time and time-stamp

Support for a precision time-stamp was introduced at software version 1.85. Calendar time in the header is now UTC (universal coordinated time). Exactly how close to UTC time depends on the acquisition hardware. If timestamp_mode = 2, then the acquisition controller is phase-locked to a GPS receiver and the timestamp is very accurate (within 10us). This feature supports multiple site data acquisition.

4 MIRF file structure

A MIRF record (or file) consists of a variably sized header block followed by de-multiplexed data channels. The data channels are normally encoded in the native format of the acquisition system with which they are recorded. A format code in the header instructs translation software how to interpret the data.

All of the information in the header block is represented in 32-bit values, mainly as signed integers but also as 32-bit IEEE754 floating point where necessary. The 32-bit values are stored in the PC compatible little-endian byte order.

Unless stated otherwise, all depths, offsets and distances are measured in the same system of units, either millimetres or millifeet as specified by the "measurement units" field.

The header block is split into two sub-blocks. The first block is always 512 bytes in size and contains general information which applies to all channels. The size of the second block depends upon the number of channels defined within the record.

The first block is known as the "general header block". One of the fields in the general header block specifies the number of channels defined in the record. This information is used to determine the size of the second block, known as the "channel header block" and which contains an array of contiguous channel structures, requiring 64 bytes per channel.

Please note that the number of channels defined within the header is not necessarily the same as the number of channels for which data actually exists. Each channel may have a different number of samples recorded and in particular a channel which has zero samples (NS = 0) should be considered to be "switched off".

5 Channel attributes

The channel attributes are stored in the channel header block and provide information specific to a particular channel.

Each channel is assigned an "owner" and possibly also a "descriptor" attribute. The owner attribute tells us to which receiver the channel belongs and the descriptor attribute tells us the signal component.

Other attributes include dc offset and voltage scaling factors which are required for the signal decoding algorithm.

6 General Header Block

Integer	Name	Description/Possible values
1	MIRF_version	7
2	File_type	0 = raw data 1 = stacked data
3	Format_code	-1 = format varies (get format from channel header block) 0 = DAQ IFP (obsolete) 1 = Multilock IFP (obsolete) 2 = Geochain IFP 3 = 32-bit integer 4 = 32-bit IEEE754 floating point 5 = 24-bit integer 6 = 16-bit integer 7 = 16-bit Delta IFP (obsolete)
4	Correlation_flag	0 = raw data 1 = correlated data
5	Controller_type	10 = DAQ (obsolete) 20 = MAU (obsolete) 30 = ASP (obsolete) 40 = GSP 50 = TSP (obsolete) 60 = SIMULATOR 70 = MONITOR
6	Tool_system	0 = ANALOG 1 = MULTILOCK (obsolete) 2 = SST500 (obsolete) 3 = GEOCHAIN 4 = MSR (obsolete) 5 = SMWD (obsolete) 6 = DELTA (obsolete) 7 = RWSPSM (obsolete)
7	Channels_defined	Determines size of channel header block

8	Test_mode	0 = TEST OFF (Normal) generic test modes -1 = IMPULSE_TEST 1 = SINEWAVE 2 = DAQ_DISTORTION_LO (obsolete) 3 = NOISE_TEST 4 = DAQ_PREGAIN (obsolete) VSP Test Modes -2 = PULSE_ALL -3 = PULSE_VZ -4 = PULSE_HX -5 = PULSE_HY -6 = PULSE_DH -7 = EPULSE -8 = TAS_SINE -9 = SIM_SINE -10 = TAS_ZEROES Geochain telemetry test modes -30 = LRX_UNEQ_LINE_RESPONSE -31 = LRX_UNEQ_RANDOM_DATA -32 = LRX_EQ_DATA_QUALITY -33 = LRX_MTX_LOOPBACK_TEST Formatted Test Modes 20=DIGI SINE BURST 21= DIGI SINE 22=SHIFT BIT
9	Dataset_id	Used to group related data. For a simple survey, dataset_id = source_id, by default. If same source is used on multiple runs, then records from each run are normally grouped by dataset_id.
10	Year	UTC year at first sample.
11	Month	UTC month at first sample.
12	Day	UTC day at first sample.
13	Hour	UTC hour at first sample.
14	Minute	UTC minute at first sample.
15	Second	UTC second at first sample.
16	Timezone_bias_seconds	Number of seconds by which local time differs from UTC time.
17	Source_id	Source number. Normally zero for a system test record.
18	Number_of_receivers	The number of receivers is supplied by the user. It should be interpreted as the maximum number of receivers for which data may be present.
19	Receiver_number	Obsolete – used for early single-receiver stack files.
20	Number_in_stack	Number of records used to create a stack file.
21	Measurement_units	1 = millimetres 2 = millifeet All elevation, depth and distance values in the header are represented in these units, unless otherwise stated.
22	Receiver_polarity	Receiver polarity for the tool system, as defined in the sensor configuration file. This is currently only meaningful for a vertical sensor. 0 = undefined 1 = SEG normal -1 = SEG reverse

23	Source_reference_channel	Multiple channels may be assigned to the REF descriptor. This field identifies the particular channel holding the reference signal applicable to this particular record.
24	reserved	
25	Record_number	Record number is unique within a survey and forms part of the record filename. Record numbers are allocated sequentially.
26	Stack_number	Records with the same stack number may be stacked with each other.
27	Fix_number	Positioning system shot fix identifier. May be used to merge missing or revised source coordinates into a dataset. Operators often strive to keep fix number the same as the record number, but this is neither guaranteed nor required.
28	Tool_MD	Measured depth at the tool reference. The tool reference is the point on the tool string that is positioned at the well reference elevation at the time that the depth measurement is zeroed. For safety reasons, the tool reference is normally either at the top tool or at the telemetry package.
29	SIus	Sample interval, measured in microseconds. To properly interpret a Geochain telemetry test record, the following substitutions must be performed. For 1us, read 0.9765625us For 2us, read 1.953125us For 4us, read 3.90625us.
30	Line_number	Not used. Dataset_id is the preferred method of grouping records.
31	Gun_pressure	Airgun firing pressure measured in psi.
32	Software_version_x100	The ACQ software version used to acquire the record, multiplied by 100. E.g. software version 1.85 is stored as the integer value 185.
33	SCX	Source easting coordinate relative to wellhead.
34	SCY	Source northing coordinate relative to wellhead.
35	TCX	Target source easting coordinate relative to wellhead.
36	TCY	Target source northing coordinate relative to wellhead. Note that TCX and TCY now normally contain the same values as SCX and SCY, so may be ignored.
37	WRE (aka WREF)	Well reference elevation, also known as “logging datum” or “drilling datum”. The elevation of the well’s depth reference point with respect to the permanent datum. Usually defined either as the top of the KB (Kelley bushing) or DF (drill floor). Typically, the elevation of the top of the KB is identical to the elevation of the DF.
38	SRE (aka SREF)	Source reference elevation. Airgun depths are measured with respect to the current SRE. SRE is normally zero for a marine survey. For a land survey over rough terrain, the elevation of ground level at one source location may be quite different to the elevation of ground level at another location. A pre-surveyed peg marking SRE provides a convenient local reference against which a vibrator pad elevation or an airgun depth, may be measured.
39	SD	Source depth below SRE.
40	S2M	Distance from source to source monitor transducer.
41	Source_elevation_error	In a heaving sea, the actual source elevation may vary from shot to shot. An integrating source motion sensor may be used to determine the error in elevation for each shot.
42	External_reference_delay_us	Typically the source controller radio delay, measured in microseconds. The reference channel must be advanced by this delay (transit time increases with positive delay).
43	Supplied_Ts_us	Time correction to seismic datum, measured in microseconds.
44	TB_advance_us	Delay from TB (electronic timebreak) to first recorded sample. This field provided because TB may occur between sampling instants.

45	Tool_skew_us	Timing skew correction from surface to digital downhole channels in microseconds, as already applied internally by the acquisition controller (diagnostic).
46	Raw_SCX_x10	The raw source easting coordinate as received in a positioning fix string, no units conversion, but multiplied by 10 (diagnostic).
47	Raw_SCY_x10	The raw source northing coordinate as received in a positioning fix string, no units conversion, but multiplied by 10 (diagnostic).
48	Controller_second	Modulo N seconds tick count. Controller ticks provide short term holdover timing continuity in the event of GPS signal loss.
49	Microsecond	Elapsed microseconds from the last controller tick that occurred before the first sample. Holds precision UTC microseconds if timestamp_mode = 2.
50	Timestamp_mode	Specifies source of the timestamp. 0 = local time from PC 1 = UTC time from PC 2 = UTC time from GPS (accurate to within 10us of UTC).
51	SDE	Seismic datum elevation.
52	Error_control	0 = None 1 = Parity Control
53	Microseismic_mode	0 = Non-Microseismic Data 1 = Microseismic Data
54	Overlap_samples	Number of samples to overlap in consecutive records (typically 0 for non-microseismic files, 1 for microseismic)
55-81	GPRMC_data	A copy of the last GPRMC GPS message received
82-128	Reserved	

7 Channel Header Block

The channel header block contains a contiguous array of channel structures. The number of channel structures is specified in the "channels_defined" variable in the general header block. The first channel structure contains information for channel one, etc. Each channel structure comprises sixteen 32-bit little-endian fields (64 bytes per channel). The 32-bit fields to be interpreted as either 32-bit signed integers or 32-bit IEEE754 floating point, as indicated below.

Note that a channel structure will exist for all channels available to the instrument, but not all of these channels will have been used. The NS field specifies the number of samples present for any particular channel, a value of zero signalling that the channel was switched off and so has no data. Use the format_code to determine the number of bytes required for each sample.

8 Channel structure

Field	Type	Attribute	Description
1	Int32	Legacy_Owner	This field maintained for compatibility with legacy MIRF readers. New readers should use field 6 below. 0 = Reference channel 1-98 = Receiver 1-98 99 = Auxiliary channel
2	Int32	Descriptor	0 = no descriptor 1 = VZ (vertical component) 2 = HX (horizontal transverse component) 3 = HY (horizontal axial component) 4 = DH (downhole hydrophone) 5 = HYDRO (monitor hydrophone) 6 = PILOT (Vibroseis pilot sweep) 7 = TB (timebreak) 8 = GF (Vibroseis ground force) 9 = RM (Vibroseis reaction mass) 10 = BP (Vibroseis base-plate) 11 = PPS (GPS Pulse Per Second) 12 = VSIM (Vibrator similarity) 13 = RSIM (Radio similarity) 14 = NGEO (Near-field Geophone) 15 = FGEO (Far-field Geophone)
3	Int32	Format_code	Data format for this channel
4	Int32	NS	Number of samples recorded for this channel. Interpret NS=0 to mean that this channel unused.
5	Int32	Pointer_us	Cursor position in microseconds.
6	Int32	Owner	-2 = AUX -1 = REF 0 = none 1...N = Receiver 1... Receiver N
7	Int32	RCX	Receiver easting coordinate relative to wellhead.
8	Int32	RCY	Receiver northing coordinate relative to wellhead.
9	Int32	TVD	True vertical depth below WRE.
10	Int32	MDO	Measured depth offset from tool reference. Add to Tool_MD to obtain the correct measured depth for this channel.
11	Float32	HSI	See Section 13
12	Int32	Reserved	
13	Float32	SSF	Per-channel sensor scaling factor. Used to apply client-defined polarity or amplitude correction. Note: this value is (integer) zero for pre-MIRF4 files.
14	Float32	DC	DC offset in volts.
15	Float32	SF	Scaling factor to convert sample value to input voltage. Note: this value is (integer) zero for original MIRF-1 files.
16	Float32	Max_magnitude	Greatest sample magnitude present in this channel

9 Sample Encoding formats

A MIRF file contains a variably sized header block followed by contiguous de-multiplexed channel data blocks arranged in channel order. The size of each channel block is determined by the number of samples recorded for that channel multiplied by the number of bytes required to encode each sample. The number of samples recorded for a given channel is specified by the NS field of the channel structure for that channel and may be zero. The number of bytes required for each sample depends upon the format_code, which specifies how samples are encoded.

To facilitate data transmission and minimise storage requirements, raw data is often encoded to sixteen bits (two bytes) per sample using the instantaneous floating point (IFP) format native to the acquisition hardware. There is currently only one IFP format in use: Geochain (also used by the unsupported SST500 and MSR systems). There are also three obsolete IFP formats defined: DAQ, Multilock and Delta. The decoding algorithm is similar for all IFP formats.

The GSP controller uses an efficient 24-bit integer format (three bytes) for surface channels. Two 32-bit (four byte) encoding formats are also defined, 32-bit signed integer and 32-bit IEEE754 floating point. The floating point format is always used for pre-processed data (e.g. correlated or stacked data).

A signed 16-bit integer format (2 bytes) is also used by certain downhole systems.

9.1 Format –1 (FORMAT_VARIES)

From MIRF version 3 onwards, each trace may have a different data format. If the format_code value in the general header block equals –1 (0xffffffff), then use the per-channel format_code from the channel header block to determine the correct data format.

9.2 Format 0 (DAQ IFP) (obsolete)

Format 0 uses sixteen bits (two bytes) per sample. The fourteen most significant bits hold the twos complement mantissa and the two least significant bits hold the exponent. The exponent may have one of four possible values, zero through three, corresponding to the four possible gains of x1, x8, x64 and x512.

9.3 Format 1 (Multilock IFP) (obsolete)

Format 1 uses sixteen bits (two bytes) per sample. The twelve most significant bits hold the twos complement mantissa and the four least significant bits hold the exponent. The exponent may have one of twelve possible values, zero through eleven, corresponding to the twelve possible instrument gains of x1, x2, x4, x8, x16, x32, x64, x128, x256, x512, x1024 and x2048.

9.4 Format 2 (Geochain IFP, also used for SST500 & MSR)

Format 2 uses sixteen bits (two bytes) per sample. The fourteen most significant bits hold the twos complement mantissa and the two least significant bits hold the exponent. The exponent may have one of four possible values, zero through three, corresponding to the four possible instrument gains of x1, x4, x16 and x64.

The last MIRF file in a Continuous Recording session will be padded out with zeroes to ensure it is 1 seconds in duration.

9.5 IFP Error handling

A consequence of the Geochain IFP coding algorithm is that the numbers 0, 1 & 2 are not valid codes. For example, a sample value of zero would cause both exponent bits to be set (maximum instrument gain), yielding a coded integer value 3. The integers 0, 1 & 2 are therefore reserved to indicate sample status as follows:

0 = no data received.
1 = detected telemetry error
2 = reserved.

Any sample flagged as a detected telemetry error, should normally be replaced by a value interpolated from adjacent good samples. Some caution is required, if many contiguous errors are present, then the record may be unacceptably distorted.

9.6 Format 3 (32-bit integer)

Format 3 uses thirty-two bits (four bytes) per sample, representing a standard 32-bit binary signed integer value. This format is defined for possible future use.

9.7 Format 4 (32-bit floating point)

Format 4 uses thirty-two bits (four bytes) per sample, representing a little-endian IEEE754 floating point value, the standard "float" type in the "C" language. This format is used for stack files generated by the ACQ acquisition software and by AS-272 digitisers in formatted mode.

For AS-272 digitisers, the full-scale value is `0.0390625V.`
Error conditions are indicated by out-of-range values.
A value of 1.0 indicates a receiver missing sample, the equivalent of ‘0’, NO_DATA in IFP mode.
A value of 2.0 indicates a telemetry missing sample, the equivalent of ‘1’, TELEMETRY_ERROR in IFP mode.

9.8 Format 5 (24-bit Integer)

Format 5 uses twenty four bits (three bytes) per sample, representing a 24-bit binary signed integer value. This format is big-endian, the first byte holds the eight most significant bits. This format is used for GSP surface channels.

9.9 Format 6 (16-bit Integer)

Format 6 uses sixteen bits (two bytes) per sample, representing a standard little-endian 16-bit binary signed integer value, the standard “short” type in the “C” language. This format is used by SMWD systems.

9.10 Format 7 (Delta IFP) (obsolete)

Format 7 uses sixteen bits (two bytes) per sample. The twelve least significant bits hold the twos complement mantissa and the four most significant bits hold the exponent. The exponent may have one of twelve possible values, zero through eleven, corresponding to the twelve possible instrument gains of x1, x2, x4, x8, x16, x32, x64, x128, x256, x512, x1024 and x2048.

10 Sample decoding

The decoding process is very similar for all three IFP formats. The differences are the size of the mantissa and the number and meaning of the exponent gain codes. The following explanation uses 32-bit little-endian "C" language semantics. Avalon Sciences can supply a copy of the "C" source file which contains our own decoding algorithms.

In the case of the Geochain format (format 2) the decoding procedure is as follows.

(1) Access the two-byte IFP sample as a 16-bit "short".
(2) Check that sample is not one of the special flag values 0,1 or 2.
(3) Use the two least significant bits to obtain the exponent integer (gain code, 0-3).
(4) Divide the sample integer by four (two right arithmetic bit-shifts) to obtain the mantissa integer.
(5) Convert the mantissa to floating-point form.
(6) Divide the mantissa by the IFP gain value obtained from the gain code exponent.
(7) Multiply the result by the per-channel scaling factor (SF), to recover instrument input terminal voltage.
(8) Subtract the per-channel DC_volts to remove the measured dc offset.
(9) Multiply by the per-channel sensor_scaling_factor (SSF) to apply any sensor polarity or calibration correction.

11 Legacy MIRF exceptions

The last three steps (scaling to input voltage, dc offset subtraction and sensor calibration) should be applied to all formats. Note however that for pre-MIRF4 files, either or both of the SF or sensor_SF fields may contain (integer) zero, so must be ignored. The recommended procedure is to check if the value is equal to zero and if so then replace the zero with unity.

12 High Side Indication

This field (field 11) is only used for channels representing a three axis receiver that supports High Side Indication. When used, this field represents a different attribute for each of the three channels (as determined by the Channel Descriptor field – 2) in a receiver as follows:

Channel Descriptor	Type	Attribute	Value
VZ - Vertical component	Bool	HSI Valid	0 = NOT valid, 1 = Valid
HX – Horizontal Transverse component	Float32	HSI Inclination	HSI Inclination Angle (0 to +180 degrees)
HY – Horizontal Axial component	Float32	HSI Roll	HSI Roll Angle (0 to +360 degrees)