

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HAYSTACK OBSERVATORY

WESTFORD, MASSACHUSETTS 01886

Telephone: 978-692-4764

Fax: 781-981-0590

10 December 2009

TO: Mark 5 group
FROM: Alan Whitney, Walter Briskin (NRAO), Chet Ruszczyk
SUBJECT: Specification for enhanced Mark 5 module directory

General

This memo details a proposed updated module-directory structure to be adopted by Mark 5C, and optionally by Mark 5A and Mark 5B and supersedes memo 047. The primary goals of this reorganization are:

1. Add sufficient additional information that DirList can give a complete summary of each scan without reference to the data in the scan.
2. Move the Disk Module Status (DMS) information out of the VSN area and into a directory-header area, along with directory-version and VSN information.

Directory structure

The proposed directory structure consists of a 128-byte directory header, followed by the scan directory. Each scan-directory entry is also a 128-byte structure. The 128-byte length aligns with disk sectors and is easy to read in a hex dump. A directory space of 10MB is available on the Mark 5 disk module, which is enough to accommodate over 75,000 128-byte scan entries, which should be enough for all present and future needs.

Directory header

The 128-byte directory header, shown in Table 1, is the first 128-byte block in the directory; the contents of the directory header is independent of the data type.

The directory header is initialized on a 'reset=erase' or 'reset=nberase' (non-bank mode erase) command. Storing the VSN in this structure gives both quick access to the VSN information as well as some protection against accidental VSN erasure. The 'VSN of non-bank-mode companion module' will be valuable during non-bank operation (e.g. >2Gbps recording) when, under the current operation of Conduant software, the VSN of the module in Bank B is obliterated; the VSN of the companion module would, of course, have to be captured before the execution of the 'reset=nberase' command actually places the modules into non-bank mode.

Name	Length (bytes)	Type	Comments
directory version#	4	integer	Version shown in this memo is Version '1'
status	4	binary	3 bits for Disk Module Status plus 1 bit for bank/non-bank mode
module VSN	32	ASCII	module VSN
companion/'continued from' module VSN	32	ASCII	Bank mode: VSN of module from which first scan is continued (if any) Non-bank mode: VSN of non-bank-mode companion module (in Bank B)
'continued to' module VSN	32	ASCII	Bank mode only: VSN of module on which last scan is continued
spare	24	integer	Set to all zero

Table 1: Contents of 128-byte directory header

Scan-directory entry

A 128-byte scan-directory entry is created referencing the data recorded each 'record=on'/'record=off' command pair. Each scan-directory entry is divided into two contiguous parts:

- Part 1: 64 bytes of information common to all data types, as shown in Table 2.
- Part 2: 64 bytes of information particular to the data type
 - Table 3 shows the format for Mark 5A/Mark 5B data types
 - Table 4 shows the format for VDIF data type

Name	Length (bytes)	Type	Comments
data type	1	integer	0 empty directory entry (end of directory) 1 unknown 2 SS test-pattern (no frame headers) 3 Mark 5A: VLBA-format data (parity stripped) 4 Mark 5A: Mark4-format data (parity stripped) 5 Mark 5A: 'Straight-through' VLBA-format data 6 Mark 5A: 'Straight-through' Mark 4-format data 8 Mark 5B: Normal data 9 Mark 5B emulation written on Mark 5C 10 VDIF format
scan number	3	integer	scan serial # on Mark 5 module (starts at 1); (upper 2 or 3 bits to be used as flags to indicate 'continued from', 'continued to', or abnormally terminated scans)
frame length	2	integer	Relevant only for VDIF format on Mark 5C: Data Frame length (bytes), including header
station code	2	ASCII	Mk5A/B: station code; 2 chars max Mk5C: unassigned; all zeroes
scan name	32	ASCII	scan name; first 31 chars used for user-specified scan name; last character reserved for software-generated suffix (if no such suffix, last character is null)
experiment name	8	ASCII	experiment name; 8 chars max
start byte number	8	integer	start-byte number of scan
stop byte number	8	integer	stop-byte number of scan

Table 2: Contents of first 64 bytes of scan-directory entry (all data types)

first time-tag	8	BCD/ integer	integer-second from first recorded data-frame header: <u>Mark 5A</u> : Mark 4 headers: 'ydoymmss' (BCD) VLBA headers: 'jjjhmmss' (BCD) <u>Mark 5B</u> : 'yyyydoyhmmss' (BCD); year is resolved with assistance of year-offset parameter in disk-frame header; not relevant for SS test-pattern data or Mark 5A internal-tvg since no data headers are present.
first frame #	4	integer	<u>Mark 5A or 5B</u> : frame number corresponding to first recorded frame header (modulo one second); if fill pattern, same rule as time tag; not relevant for SS test-pattern or Mark 5A internal-tvg data.
byte offset	4	integer	<u>Mark 5A or 5B</u> : byte offset from first recorded byte to beginning of first frame header; if fill pattern, same rule as time tag; guaranteed to be reliable if previous two entries are filled since they can't be filled without knowing the byte offset. Not relevant for SS test-pattern or Mark 5A internal-tvg data.
track data rate/ bit-stream data rate	4	integer	<u>Mk5A</u> : track data rate (Mbps); Not relevant for SS test- pattern data or internal tvg. <u>Mk5B</u> : bit-stream data rate (Mbps); not relevant for SS test pattern data or internal tvg
#tracks/ bit-stream mask	4	int/binary	<u>Mk 5A</u> : #tracks; 32 if 'st' mode <u>Mk 5B</u> : bit-stream mask (also for Mk5C emulation of Mk5B)
unassigned	40	-	Set to all '0'

Table 3: Contents of last 64 bytes of a single scan-directory entry for Mark 5A/Mark 5B data formats

Data Group 1	2	integer	#chans/thread ₆	Base Thread ID ₁₀
	2	integer	C ₁ #bits/sample ₅	#threads ₁₀
	2	integer	samples/sec/chan/thread (units of 125 kbps)	
	2	integer	Station Code ₁₆	
Data Group 2	2	integer	#chans/thread ₆	Base Thread ID ₁₀
	2	integer	C ₁ #bits/sample ₅	#threads ₁₀
	2	integer	data-rate/thread (units of 125kHz) ₁₆	
	2	integer	Station Code ₁₆	
.....(more Data Groups, as necessary; 8 max)	48			

Table 4: Contents of last 64 bytes of a single scan-directory entry for VDIF data format;
subscript indicates the # of bits in the corresponding parameter

where #chans/thread, Base Thread ID, 'C', #bits/sample and Station Code are all taken from a VDIF Data Frame Header from the corresponding Data Group (see VDIF specification); data-rate/thread is specified in units of 125 kbps.

When using the VDIF format, note that Mark 5C imposes the following restrictions: 1) all Data Frames within a scan must have the same length, 2) only a single Station ID is allowed within each 'simple' Data Stream and within each Data Group of a 'complex' Data Stream.

Though there is no prohibition on mixing scans with different data types on a single module (or non-bank module-pair), the user should verify that the target processor can properly interpret such data.

DirList

The DirList listing should be augmented to display additional information residing in the updated directory structure.

Mark 5A

It is suggested that DirList have an option to specify the year so that the year ambiguity can be removed from the DirList output; otherwise DirList will choose the closest past date. A sample DirList listing might be (bank-mode example, shown data continued from/to other modules):

```
VSN: XYZ-0012/2000/1024
(continued from VSN: XYZ-0011/2000/1024)          (continued to VSN:XYZ-0013/2000/1024)
545 scans      100% full (2000/2000GB)
Exp: grf051    Stn: ef          Data format: Mark 5A
```

Scan #	Scan name	Start time	Length(sec)	Mode	TrackRate	Total Rate
1-	223-0056	2006:223:00:56:05.000	300	vlba:16	16	256
....						
204	223-0815	2006:223:08:15:00.000	30	vlba:32	16	512
....						
545+	223-2110	2006:223:21:10:03.000	150	vlba:32	16	512

The ‘Status’ column is a single-character flag that indicates the scan status:

- blank indicates OK
- scan continued from another module
- + scan continue to another module
- x scan terminated abnormally or other fault

Mark 5B

An example of Mark 5B DirList listing might be (non-bank mode example):

```
VSN: XYZ-0012/2000/1024      (companion VSN: XYZ-0016/2000/1024)
545 scans      83% full (3320/4000GB)
Exp: grf051    Stn: ef          Data format: Mark 5B
```

Scan #	Scan name	Start time	Length (sec)	bit-mask	bit-stream rate (Mps)	Total Rate (Mbps)
1	123-0056	2006:123:00:56:05.000	300	55555555	64	1024
2	123-0110	2006:123:01:10:03.000	150	f f f f f f f f	64	2048
....						

If multiple experiments are present on the module, the ‘Exp/Stn/Data format’ line is re-printed before the corresponding set of scans. If the bit-stream mask is unavailable for some reason (such as it wasn’t specified when data were transferred over a network), then the ‘bit-mask’ and ‘Sample Rate’ cannot be determined and the associated table entries will be ‘?’.

VDIF

An example of VDIF DirList listing might be (non-bank mode example):

```
VSN: XYZ-0012/2000/1024      (companion VSN: XYZ-0016/2000/1024)
545 scans      83% full (3320/8000GB)
Exp: grf051      Stn: ef      Data format: VDIF
Scan  Scanname      Stn      Start time      Data span      Format      Frame length
#                                           (sec)                                           (bytes)
  1    123-0056      ef      2006:123:00:56:05    300      1024-16-2-16    8016
 2*   123-0110      ef      2006:123:01:10:03    150      1024-16-2-16    8016
 2*   123-0110      eg      2006:123:01:10:03    150      256-8-2-1       8016
.....
```

where Scan 1 is a 'simple' VDIF scan, and Scan 2 shows two Data Groups within a single 'complex' VDIF scan (indicated by '*'; see VDIF specification for further explanation).

Notes

1. Note that the suggested DirList information comes only from the directory without reference to the actual data, which dramatically improves the speed at which the DirList can be created.
2. If multiple experiments reside on the same disk module, a new 'Exp:....Stn:....' heading will be issued preceding the affected scans. Unknown experiment names and station codes will be listed as 'EXP' and 'STN' respectively.
3. An attempt to record a scan with a duplicate scan name will cause a trailing alphabetical character ('a-z', then 'A-Z') to be automatically appended to the scan name. If there are more than 52 scans with same user-specified name, the suffix sequence will repeat.
4. When all or part of a scan is copied to a Linux file (using disk2file), the default assigned file name is:
Mark 5A data format: '<scan label>.mk5a'
Mark 5B data format: '<scan label>_bm=<bit mask>.mk5b' (also for Mk5C emulation of Mk5B)
VDIF data format: '<scan label>_fd=<VDIF Format Designator>.vdif'
5. If the experiment name in the scan-directory entry is null, it will be replaced with 'EXP' in the filename. Similarly, a null station code will be replaced with 'STN' in the filename.
6. When Mark 5 data are copied to a disk module (file2disk), the experiment name, station code, scan name and bit-stream mask (Mark 5B only) are parsed from the file name; the data in the file are examined to determine the remainder of the parameters to be written to the associated scan-directory entry.

Addendum: Scan names, Scan Labels and Linux filenames

Mark5 defines a ‘scan’ as the set of recorded data between a ‘record=on’ and ‘record=off’ command pair. Each scan is identified by a scan name, experiment name and station code, which are normally derived from the information in the associated VEX file used in the scheduling of the experiment (see <http://lupus.gsfc.nasa.gov/vex/vex.html>). An attempt to record a scan with a duplicate scan name on the same disk module will cause a trailing alphabetical character (‘a-z’, then ‘A-Z’) to be automatically appended to the scan name. If there are more than 52 scans with same user-specified name, the suffix sequence will repeat. Information about the experiment name, station code, bit-stream mask, and sample rate are stored in the associated directory entry.

A scan label is defined as the character string

`<exp name>_<stn code>_<scan name>`

where

`<exp name>` is the name of the experiment (e.g. ‘grf103’); maximum 8 characters, but by convention corresponds to a standardized 6-character experiment name. If null, will be replaced with ‘EXP’.

`<stn code>` is the station code (e.g. ‘ef’); maximum 8 characters, but by convention corresponds to standardized 2-character codes. If null, will be replaced with ‘STN’

`<scan name>`¹ is the identifier for the scan (e.g. ‘254-1056’), usually assigned by the observation-scheduling program; max 31 characters, though may be augmented to 32 characters by automatically generated duplicate-breaking suffix character.

Maximum scan-label length, including embedded underscores and possible scan-name suffix character, is 50 characters. `<experiment name>`, `<station code>` and `<scan name>` may contain only standard alpha-numeric characters, except ‘+’ and ‘-’ characters may also be used in `<scan name>`. All fields are case sensitive. No white space is allowed in any of these subfields.

Lower-case characters in all subfields are preferred. An example scan label is:

grf103_ef_scan001

When a Mark 5B scan (or portion of a scan) is copied to a Linux file with *disk2file*, a Linux filename compatible with the internationally agreed e-VLBI filenaming convention (reference <http://www.haystack.edu/tech/vlbi/evlbi/memo.html> memo #49) is assigned as

‘<scan label>_bm=<bit-stream mask>.mk5b’
(example: ‘grf103_ef_scan001_bm=0x0000ffff.m5b’)

Linux files to be transferred to a Mark 5B disk via the *file2disk* should have filenames corresponding to the standardized format described above so that the associated Mark 5B directory entries can be properly filled.

Similarly, the Linux filename for a VDIF-format file should conform to the VDIF file-naming convention (see VDIF specification) as, for example:

‘gre53_ef_scan035_fd=1024-16-2-16.vdif’

¹ The `<scan name>` described here is equivalent to what is called `<scan_ID>` in VEX files, except the set of legal characters in `<scan name>` is more restrictive and must be observed.

which specifies data from experiment 'gre53', station 'ef', scan name 'scan035' with the VDIF Format Designator '1024-16-2-16'.