MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HAYSTACK OBSERVATORY

WESTFORD, MASSACHUSETTS 01886

18 August 2006

TO: Mark 5 group

FROM: Alan Whitney, Roger Cappallo, Kevin Dudevoir, Ed Himwich, Dan Smythe,
                 Mike Titus

SUBJECT: Specifications for enhanced Mark 5 module directory

General

This memo details proposed changes to the module-directory structure of the Mark 5B and, possibly, the Mark 5A module-directory. 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.

3. Maintain backwards compatibility to the greatest degree possible.

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 is the first 128-byte block in the directory and contains the following information:

Name	Length (bytes)	Type	Comments
directory version	4	integer	propose to start at '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 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	initialize to a zero

Table 1: Contents of directory header

The directory header will be 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 (and which Conduant does not seem anxious to change), 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.

Scan-directory entry

Note on Mark 5B: The upper 4-bits of the (currently) 16-bit 'User-specified' data field in the disk-frame header will be co-opted to store the year offset from 2000. This allows the year and DOY to be determined unambiguously on any Mark 5B recording through 2021 (since it is known that no existing Mark 5B recordings were made before 2006), which is likely beyond the life of the Mark 5B. The user data field in the disk-frame header will shrink to 12 bits.

Each scan-directory entry includes the following information, as follows:

Name	Length (bytes)	Type	Comments
data type	4	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 7 Mark 5A: tvg data (no frame headers) 8 Mark 5B: Normal data 9 Mark 5B: tvg data 10 Mark 5B: numerical-ramp data
scan number	4	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)
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

station code	8	ASCII	station code; 8 chars max
start byte number	8	integer	start-byte number of scan
stop byte number	8	integer	stop-byte number of scan

first time-tag	8	BCD	time tag of first tape(5A)/disk(5B) data-frame header to integer-second precision:
			Mark 5A
			   Mark 4 headers: 'ydoymmss'
			   VLBA headers: 'jjjhhmmss'
			Mark 5B: 'yyyydoyhhmmss' (year is resolved with assistance
			   of new year-offset parameter in disk-frame header).
			If first frame header position is fill pattern, time tag of frame header that would have been in that position. Not relevant for SS test-pattern data or Mark 5A internal-tvg since no data headers are present.

first header #	4	integer	frame-header number corresponding to first frame header (defined as zero on second tick, and explicitly written into Mark 5B frame header); if fill pattern, same rule as time tag. For Mark 5B; will always be zero for locally recorded scan. Not relevant for SS test-pattern or Mark 5A internal-tvg data.
byte offset	4	integer	byte offset 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.
scan length	4	integer	# of complete data frames, including all fill pattern, if any; not relevant for SS test-pattern or Mark 5A internal-tvg data.
total data rate	4	integer	total data rate (Mbps); Not relevant for SS test-pattern data or Mark 5A internal-tvg.
#tracks/bit-stream mask	4	int/binary	Mark 5A: #tracks; 32 if 'st' mode Mark 5B: bit-stream mask
spare	28	-	Initialized to all zero's

Table 2: Contents of one scan-directory entry

Notes:

1. <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.

2. A Mark 5 (either 5A or 5B) <u>scan label</u> is *defined* as

   '<exp name>_<station code>_<scan name>'
   (Example: 'grf01_ef_123-0056')

3. When all or part of a scan is copied to a Linux file (using *disk2file*), the *default* assigned file name is

   Mark 5A: '<scan label>.m5a'
   Mark 5B: '<scan label>_bm=<bit mask>.m5b'

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.

4. When data in a Mark 5A or Mark 5B Linux file 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.

DirList

The DirList listing will be augmented to display the information in the directory structure. 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 #	Scanname	Flg	Start time	Length(sec)	bit-mask	SampleRate	Total Rate
1	123-0056	*	2006:123:00:56:05.000	300.0	0000ffff	32	512
2	123-0110	*	2006:123:01:10:03.000	150.0	55555555	32	512

…..

The 'Flg' column is a single-character flag that indicates the scan status:

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

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 'bs-mask' and 'Sample Rate' cannot be determined and the associated table entries will be '?'.

For 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 #	Scanname	Flg	Start time	Length(sec)	Mode	TrackRate	Total Rate
1	223-0056	-	2006:223:00:56:05.000	300.0	vlba:16	16	256
…..							
204	223-0815	x	2006:223:08:15:00.000	30	vlba:32	16	512
…..							
545	223-2110	+	2006:223:21:10:03.000	150.0	vlba:32	16	512

Note that this information comes only from the directory without reference to the actual data, which will dramatically improve the speed at which DirList can provide this information.

If multiple experiments or stations 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.

Other specifications and notes

1. 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.

2. Each directory entry will be defined by a structure, and the directory itself will be an array of these structures. This allows easy definition and access to individual scan elements – for example:

   struct scan_tag
   .
   .
   char scanname [32];
   .
   struct scan_tag   scans [MAXSCANS]

   Individual scan parameters are easily accessed in this structure. For example, the scanname of scan 37 may be retrieved as

   strcpy(dest,scans[37].scanname)

   etc.

3. The version of *Mark5A* that supports the features of the Mark 5A+ must be upgraded to recognize and support the new Mark 5B directory structure.

The following pages are extracted from the proposed *dimino* specification.

6. Scan names, Scan Labels and Linux filenames

Mark5 defines a 'scan' as a <u>continuously recorded set of data</u>. Each scan is identified by a <u>scan name</u>, 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 <u>scan label</u> is defined as the character string

<p style="text-align:center"><exp name>_<stn code>_<scan name></p>

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> 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 <u>scan label</u> is:

<p style="text-align:center">grf103_ef_scan001</p>

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

<p style="text-align:center">'<scan label>_bm=<bit-stream mask>.m5b'
(example: 'grf103_ef_scan001_bm=0x0000ffff.m5b')</p>

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.

Note: 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.

disk2file – Transfer data from Mark 5 to file [command list]

Command syntax: disk2file = [<destination filename>] : [<start byte#>] : [<end byte#>] : [<option>] ;

Command response: !disk2file = <return code> ;

Query syntax: disk2file? ;

Query response: !disk2file ? <return code> : <status> : <destination filename> : <start byte#> : <current byte#> : <end byte#> : <option> ;

Purpose: Transfer data between start-scan and stop-scan pointers from Mark 5 to file.

Settable parameters:

Parameter	Type	Allowed values	Default	Comments
<dest filename>	literal ASCII	no spaces allowed	See Comments	Default <dest filename> is as specified in Section 6 (i.e. '<scan label>_bm=<bit mask>.m5b'). Filename must include path if path is not default. See Note 1.
<start byte#>	int | null		See Note 1	Absolute byte#; if null, defaults to start-scan pointer. See Note 1.
<end byte#>	int | null		See Note 1	Absolute end byte#; if preceded by '+', increment from <start byte#> by specified value; if null, defaults to stop-scan ponter. See Note 2.
<option>	char	n | w | a	n	n – create file; error if existing file w –erase existing file, if any; create new file. a – create file if necessary, or append to existing file

Monitor-only parameters:

Parameter	Type	Values	Comments
<dest filename>			Destination filename (returned even if filename was defaulted in corresponding 'disk2file' command)
<status>	char	active | inactive	Current status of transfer
<current byte#>	int		Current byte number being transferred

Notes:
1. The default <dest filename> is recommended if a subsequent *file2disk* operation is to be performed since it contains the information needed to reconstruct the original Mark 5B directory entry.
2. The 'scan_set' command is a convenient way to set the <start byte#> and <stop byte#>.
3. If <start byte#> and <end byte#> are null, the range of data defined by 'scan_set' will be transferred.
4. To abort data transfer: The 'reset=abort' command may be used to abort an active disk2file data transfer. See 'reset' command for details.
5. When <status> is 'inactive', a 'disk2file?' query returns the <dest filename> of the last transferred scan, if any.

file2disk – Transfer data from file to Mark 5B

Command syntax: file2disk = <source filename> : [<start byte#>] : [<end byte#>] : [<scan label>] : [<bit-stream mask>] ;

Command response: !file2disk = <return code>;

Query syntax: file2disk? ;

Query response: !file2disk ? <return code> : <status> : <source filename> : <start byte#> : <current byte#> : <end byte#> : <scan#> : <scan label> : <bit-stream mask>;

Purpose: Initiate data transfer from file to Mark 5 data disks

Settable parameters:

Parameter	Type	Allowed values	Default	Comments
<source filename>	ASCII	no spaces allowed	'save.data' or last value	If not in standardized filename format (see Section 6), must specify at least <scan label> and recommend specifying <bit-stream mask> as well. See Note 1. Filename must include path if not default.
<start byte#>	int		0	Absolute byte number; if unspecified, assumed to be zero
<end byte#>	int		0	If =0, will copy to end of file
<scan label>	ASCII		Extracted from <source filename>	Required if <source filename> is not is standardized format (see (Section 6). See Note 1. Example: 'exp53_ef_scan123'
<bit-stream mask>	hex		0	Should be specified if <scan label> is specified. See Note 1.

Monitor-only parameters:

Parameter	Type	Values	Comments
<status>	char	active | inactive	Current status of transfer
<current byte#>	int		Current source byte# being transferred
<scan#>	int		Sequential scan number on disk module

Notes:
1. If <source filename> is in the standardized format for Mark 5B (see Section 6), *dimino* will parse the constituent fields to determine <experiment name>, <station code>, <scan name> and <bit-stream mask>. If source filename does not include a scan label in the proper format, <scan label> must be specified. If <scan label> is specified, it is recommended that <bit-stream mask> also be specified so that the Mark 5B directory entry can be properly completed.
2. The data in the source file must be in Mark 5B data format.
3. To abort data transfer: The 'reset=abort' command may be used to abort an active file2disk data transfer. See 'reset' command for details.
4. When <status> is 'inactive', a 'file2disk?' query returns <source filename> of the last transferred scan, if any.

net2disk – Transfer data from network to disks

Command syntax: net2disk = <control> : <scan label> : <bit-stream mask> ;

Command response: !net2disk = <return code> ;

Query syntax: net2disk? ;

Query response: !net2disk ? <return code> : <status> : <scan#> : <scan label> : <bit-stream mask> ;

Purpose: Enable data transfer from network to local disks

Settable parameters:

Parameter	Type	Allowed values	Default	Comments
<control>	char	open | close		'open' or 'close' socket
<scan label>	ASCII			Scan label to be assigned to this data (see Section 6); if not specified, defaults to 'EXP_STN_net2disk'
<bit-stream mask>	hex		0	<bit-stream mask> associated with data. See Note 1.

Monitor-only parameters:

Parameter	Type	Values	Comments
<status>	char	active | inactive | waiting	Current status of transfer
<scan#>	int		Sequential scan number on disk module
<scan label>	ASCII		Assigned scan label
<bit-stream mask>	hex		Assigned bit-stream mask

Notes:
1. The <bit-stream> mask should always specified so that the Mark 5B directory entry can be properly completed.
2. See Notes with 'disk2net' command for usage rules and restrictions.
3. When <status> is 'inactive', a 'net2disk?' query returns <scan label> of the last transferred scan, if any.
4. Note that the network protocol parameters are set by the 'net_protocol' command.

record – Turn recording on|off; assign scan label

Command syntax: record = <record on/off> : <scan label/name> : [<experiment name>] : [<station code>] ;

Command response: !record = <return code> ;

Query syntax: record? ;

Query response: !record ? <return code> : <status>: <scan#> : <scan label> ;

Purpose: Turn recording on|off; assign scan name, experiment name and station code

Settable parameters:

Parameter	Type	Allowed values	Default	Comments
<record on/off>	char	on	off		'on' automatically appends to the end of the existing recording. 'off' stops recording and leaves system in 'idle' mode.
<scan label/name>	ASCII	32 chars max		Relevant only if record is 'on'. If in <scan label> format, field is parsed for <exp name>, <station code> and <scan name>. Otherwise, interpreted as <scan name>, in which case <experiment name> and <station code> should be specified separately. If <scan name> is duplicate of already-recorded scan, a suffix will be added to the <scan name> part of the <scan label> -- see Note 6.
<experiment name>	ASCII	8 chars max	null	Experiment name; ignored if <record on/off> is 'off'
<station code>	ASCII	8 chars max	null	Station code; ignored if <record on/off> is 'off'

Monitor-only parameters:

Parameter	Type	Values	Comments
<status>	char	on	off	halted	throttled	overflow	waiting	'halted' indicates end-of-media was encountered while recording. 'throttled', 'overflow' and 'waiting' are all error conditions.
<scan#>	int		Sequential scan number; starts at 1 for first recorded scan.
<scan label>	ASCII		Scan label – see Notes 5 & 6. See Section 6 for definition of scan label.

Notes:
1. After record is turned 'on', the user should periodically query 'status' for details; if recording stops on its own accord (due to end-of-media, etc.), this will be reflected in the response to the 'status' query as 'recording stopped', and a 'record' query will show the status as 'halted'; a subsequent command to turn record 'off' or 'on' will reset the relevant bits (5-4) in the 'status' response.
2. When recording, the record pointer will update to show the approximate position. If the record pointer is noted not to be incrementing, an error flag is set in the 'status?' query which can be used as a first order check of proper recording.
3. When <status> is 'off', a 'record?' query returns the <scan label> of the last recorded scan, if any.

4. Typical causes for status errors:
    a. "throttled" – data rate from Mark 5B I/O card is too fast for disks to keep up (flag received by I/O board from StreamStor card)
    b. "overflow" – FIFO overflow on Mark 5B I/O card
    c. "waiting" – CLOCK has stopped or is faulty
5. The <scan label> field is created in the standardized format specified in Section 6, namely '<exp name>_<station code>_<scan name>'. If <experiment name> and/or <station code> are null, they will be replaced with 'EXP' and 'STN', respectively.
6. 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 (example: '312-1245a'). If more than 52 scans with same user-specified name, the suffix sequence will repeat.