WHITE PAPER

METADATA

METADATA: WHERE-WHAT-WHO

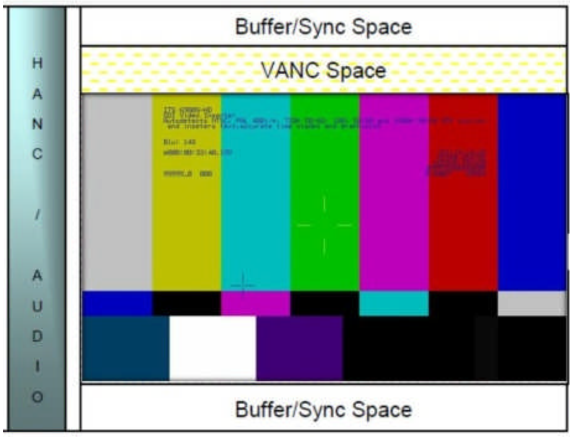
With analog video, one can embed machine readable data using edge or scan lines encoding schemes. Using scan lines, one can embed about 108 bytes of data per field/216 per frame. All of this data is impressed on vertical blanking lines. For edge encoding a few bits could be encoded at the start of each line. This approach accommodates about 12 bytes/field; 24 bytes/frame after overhead

Decoding required special equipment that could pick the data bits out of the video between the sync pulses and the image area.

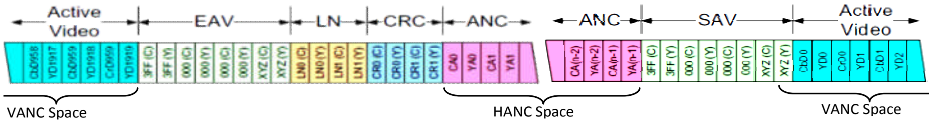
In HD-SDI video one can place more than 23K of data in a 720p/60 and more than 65K of data in a 1080p/60 video stream. In fact the data can be organized up to 100 (720/p) or 280 (1080/p) different streams of data updated at the video frame rate (29.94, 30, 59.97, 60 times/sec).

Where is Metadata

In SDI digital video SMPTE defined a significant area where data can be stored in the left side of the video between the end of active video (EAV) and of the prior frame/field and the start of active video (SAV) of the new frame/field. This area is where payload definition data (resolution, frame rate, encoding) is kept, audio (up to 16 channels of PCM encoded audio), affiliate, billing, and workflow data. This area is referred to as the HANC (Horizontal Ancillary) data area. SMPTE also defined the equivalent of a vertical blanking area where data can be placed. Vertical ancillary (VANC) data can have SAP, and other data.



When SDI video bit streams are reconstructed into images, there is a viewable image area referred to as the active video region. This region is below the VANC space (see right) up to the bottom buffer space and in between the SAV and EAV. The content of the active video area are samples of the image. Most commonly the sampling is 4:2:2 where each pixel (the gross resolution of the image, e.g. 1920x1080) is comprised of a luma value (Y) and half of a color sample (either C_r or C_b). Each sample is allocated 10 bits. When multiplexed and serialized in accordance with SMPTE 292, the stream looks like this (right).



When placing data in the HANC or VANC, the general format is still like image samples; 10 bits per word.

What is Metadata

In general an ANC packet is built with a wrapper. The wrapper has a signature pattern referred to as the Ancillary Data Flag (ADF) and is encoded by 0x000, 0x3FF, 0x3FF. This pattern is restricted to this use and may not appear in the video or the data. The signature tells the display or decoder of interest that this group of samples is not image data, but ancillary data. As said earlier, packets may contain audio data, payload data, and workflow data, film codes, SAP, closed captioning, billing and other information of import to editors, directors, broadcasters, networks and content generation people.

This information must be decoded, however, in real time and reasonably well coordinated[1] with the video associated with it.

Each packet carries in the wrapper a DID (**D**ata **ID**entification) word. These DID values indicate the type and format of data contained in the packet. This structure is all detailed in SMPTE-291M. SMPTE manages DID values and publishes them in a registry. The list of registered DID identifiers can be found at www.SMPTE-ra.org.

Things start to diverge from here. There are type 1 packets and type 2 packets. The type spec changes the structure following the DID. For our purposes and the information that MISB prescribes, type 2 packets are of most interest. In this case (type 2 packet) the DID is followed by a **S**econdary **DID** (SDID) word. As of this writing there are about 42 type 2 packets in groups (DID) that identify payload, encryption, compression, compressed audio data and more. Of interest to instrumentation and MISB are DID 44h type 2 packets. Identifier 44h is a key-length-value (KLV) data type. There are two of these packets. KLV data in the HANC space (SDID 14h) and KLV data in VANC space (SDID 04h).

MISB only uses 44h (DID) 04h (SDID) type 2 packets. These are KLV packets in VANC space as defined in SMPTE 291M.

Reading through much of the background data on the development of the MISB implementation of metadata, it becomes clear that to use the VANC space offers the key benefit of capacity. For example, in a 1080i/p format, there are 1920 samples available for data on each line. As data, this is 1920 bytes, more than eight times that available for a whole image in analog NTSC video. Multiply this by 40 VANC lines in the format one has 76.8K bytes per frame on a 1080i/p stream (2.3Mbytes/sec for 60i or 4.6 Mbytes/sec for 60p).

There is overhead associated with that (ADF, DID, SDID, DC[2], Checksum and more inside the wrapper), but this still translates to as many as 16,800 packets/second or 280 different streams of data updated on each frame. This capacity and potential data rate is a great resource for all kinds of instrumentation and telemetry purposes.

The amount of available capacity varies with format. This complicates matters a bit. The variation is in part the difference in gross resolution (Y sample count per line) and the VANC space allocated to the format. A 720p stream actually has 750 lines and only 1280 active video samples per line. Five of those lines are allocated to buffer space at the bottom, leaving 25 lines at the top. There is also the matter of line 14. This line is reserved by SMPTE as the point everyone will use to scene switch, so MISB says not to write anything there.

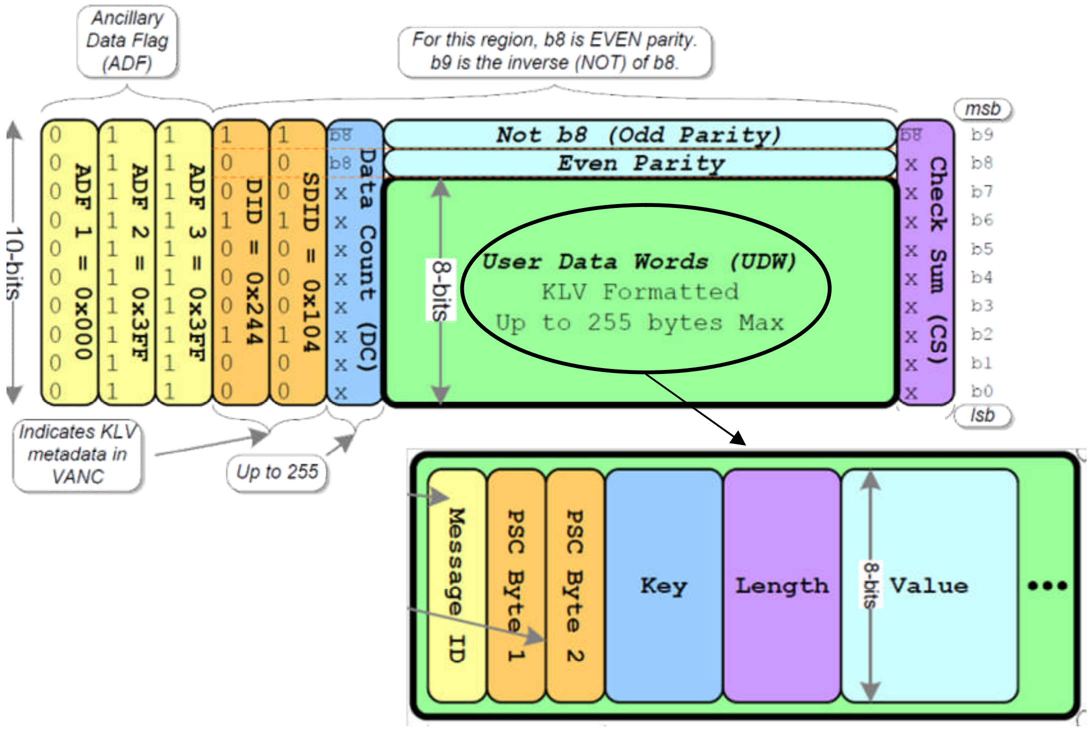
---

[1] For example, SMPTE specifies that the audio stream need only be reconstructed and realigned with the video stream to ±50ms of the related video. This is good enough to lip sync to the stream. Not so good for instrumentation.

[2] We forgot to mention the Data Count (DC) specification of a packet. This informs the decoder how many total bytes are present in the User Data Word (UDW) space to follow.

When MISB built its specifications for using KLV packets in VANC space, the foundational structure complies with the SMPTE design. This approach enables interoperability with a whole range of commercial equipment.

In accordance with SMPTE 291 and RP (SMPTE **R**ecommended **P**ractice) 214 the basic KLV packet for all MISB data is described as shown below[3].



When using a KLV packet of any type, SMPTE further defines the structure of the User Data Word (UDW, specified in SMPTE 336 and RP 214). First it is 255 words (bytes) long at most. But three bytes are consumed by the segmenting features built into the KLV structure, the Message ID (starts from decimal 1) and the **P**acket **S**equence **C**ounter (PSC, values start at 1). Then there is the KLV format itself that has 17 bytes of overhead, the key (16 bytes) and the length (1 byte). Therefore, a packet can carry up to 235 bytes of data. The table below estimates the data recording capacity and the possible number of separate 235 byte data streams per format.

Format	VANC Lines	Samples/Line (SAV-EAV)	Packs/Line	Bytes/Line
480i/60[4]	525-480-10(top/bottom)-1=29	720	2	470
720p/60	750-720-5 (bottom)-line 14=24	1280	4	940

[3] The only comment we have on this drawing is the binary data shown for the SDID word is incorrect. The hex value shown (including the parity bits) is correct, 0x104, but the binary bits shown are 0x244 which is a copy of the DID word.

[4] This is SD-SDI format and it is a bit more complicated. MISB does not define SD-SDI for KLV packet use.

Format	VANC Lines	Samples/Line (SAV-EAV)	Packs/Line	Bytes/Line
1080i/60	1125-1080-4(bottom)-line 14 = 40	1920	7	1,645[5]
1080p/60	1125-1080-4-1=40	1920	7	1,645

Who is Metadata

The value can be anything, from ASCII to binary. How one can know what the byte values represent in the value element is found in a dictionary. The dictionary is a list of keys (K of the KLV) that describes the data contained in value and how it is represented (encoded and decoded). SMPTE has a dictionary of nearly 3,000 keys[6] which they refer to as elements. MISB owns a block of keys for private use. MISB publishes the list of keys in its dictionary 807.x (807.8 as of this writing). This dictionary and the list of MISB standards can be found at http://www.gwg.nga.mil/misb/stdpubs.html.

The *Microsecond Time Pack* is defined by key 06.0E.2B.34.02.05.01.01.0E.01.01.03.11.00.00.00. That key specifies that the Microsecond Time Pack is comprised of two other data types, the *Status byte followed by an 8 byte Microsecond Time Stamp*. These are found in the dictionary by keys 06.0E.2B.34.01.01.01.01.0E.01.01.03.10.00.00.00 and 06.0E.2B.34.01.01.01.03.07.02.01.01.01.05.00.00 respectively. For example, the ***Microsecond Timestamp Status*** byte in the dictionary is fully defined as shown in the box at the right.

STATUS BYTE DEFINITION

Bit Position Value Definition

7 (MSB) 0 GPS Locked - timestamp clock locked to GPS

1 GPS Flywheel - timestamp clock not locked to GPS, so it is running on an internal oscillator

6 0 Normal - timestamp incremented normally since last message

1 Discontinuity - timestamp has not incremented normally since last message

5 0 Forward - If Bit 6=1, this indicates that the timestamp jumped forward

1 Reverse - If Bit 6=1, this indicates that the timestamp jumped backwards

4-0 (LSB) 1 Reserved Bits = 1

Each key then is a data type definition that can be used to identify the type of data found in value (V) of the KLV pack and how it is represented. The key dictionary is the master "decoder ring" and each key defines the who of the metadata.

ITS INSERTION ENGINE WRITES & READS MICROSECOND TIME PACKS

The ITS v1.0 engine can be instructed to write or read a VANC KLV packet encoded to be a Microsecond Time Pack (MTP) compliant with MISB 0605.3 and its key dictionary 807.x

MISB prescribes that an MTP be written in the VANC space (since they are type 2 KLV packets) on line 9. It also states that such packs are to be written as the first pack on line 9 in the VANC. Therefore the ITS Insertion Engine places the MTP metadata packet immediately after the SAV on line 9 of any HD-SDI format.

When instructed to read a MTP, the ITS Insertion engine looks ONLY on line 9 and expects it to be in the first metadata packet on the line after the SAV marker.

When the ITS Insertion Engine is constructed to display (human readable) the MTP, it reads it,

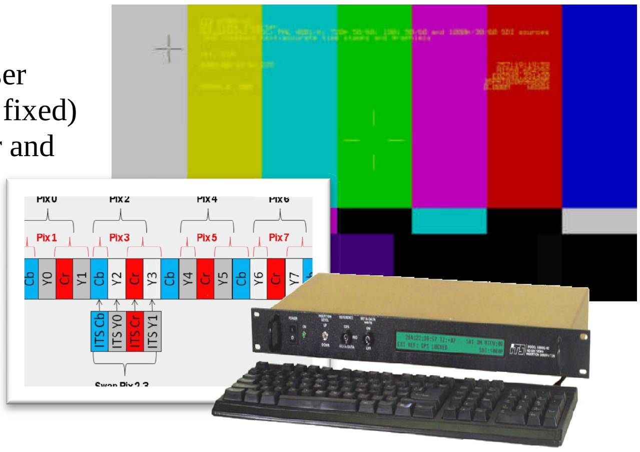
[5] In this instance the format is interlaced and, therefore, the resulting data rate is half that of 1080p/60
[6] A list of registered elements can be found at www.smpte-ra.org.

decodes it and formats it into a Gregorian or Julian human readable date and time. The time can be displayed at a resolution of 1 second to 100 µseconds at the user's option. This can be placed anywhere in the visible video area. The status of the metadata time is shown as a lower case or upper case M depending on at the time of recording the pack the time was locked (M) or unlocked (m) from a time reference (GPS or IRIG).

CODECS & RECORDERS

When selecting recorders, CODECs and when attempting to route the SDI (or encoded version of it) into a network of some type, it is important to ensure that the VANC lines are preserved in the encode process and restored in the decode process. In restoration it is also important that the time is restored to the correct field/frame of the image stream.

THE ITS HD-SDI INSERTION ENGINE



The ITS HD SDI Insertion Engine is designed overlay user ASCII human readable text and crosshairs (movable and fixed) into the HD-SDI video stream. Text can be set to a color and both the moving and fixed crosshairs can each be set to different colors.

More than 3,000 text characters can be placed in a 1080i/p image. The movable crosshair can be positioned to pixel level precision.

The HD SDI Insertion Engine employs a pixel swap technology to minimize bit stream buffering and image obscuration with the overlay text. The total process latency is less than 10 µseconds.

When integrated into one of our HD-SDI insertion products (e.g. 6980G-HD) that also employs the ITS GPS receiver, time can be collected at the equivalent of vertical sync (the first image sample of the frame or field) to an accuracy of 4±2µseconds of the GPS time mark. The time collected can be displayed anywhere on the image from a resolution of 1 second to 100 µsecond increments.

Additionally, this time is converted to POSIX time and formatted into a MISB microsecond time stamp KLV packet on each EAV event. That means time is recorded on fields and frames. The ITS HD-SDI Insertion Engine may also be set to read and make human readable previously recorded microsecond time stamps encoded into KLV packets per MISB.

For more information contact:

Instrumentation Technology Systems
(818) 886-2034 • (818) 886-7573 (fax)
phightower@itsamerica.com
www.ITSamerica.com