The ATSC Transport Layer, Including Program and System Information Protocol (PSIP)

BERNARD J. LECHNER, LIFE FELLOW, IEEE, RICHARD CHERNOCK, MEMBER, IEEE,
MARK K. EYER, SENIOR MEMBER, IEEE, ADAM GOLDBERG, SENIOR MEMBER, IEEE, AND
MATTHEW S. GOLDMAN, SENIOR MEMBER, IEEE

Invited Paper

The basic concepts of the transport layer of the ATSC digital television system, including multiplexing, timing, and synchronization, are presented. Constraints on and extensions to the MPEG-2 Systems standard are discussed, including the important extension that adds a Program and System Information Protocol (PSIP). Transport related aspects of private data carriage and conditional access are also described.

Keywords*—*Advanced Television Systems Committee (ATSC), buffer model, conditional access (CA), content advisory, digital television (DTV), directed channel change (DCC), electronic program guide (EPG), MPEG, mutiplexing, naming, navigating, numbering, packetizing, private data, Program and System Information Protocol (PSIP), sections, synchronization, tables, timing, transport layer, transport stream (TS), virtual channels.

I. INTRODUCTION

The ATSC digital television (DTV) system, as described in the ATSC DTV standard [1], provides a flexible framework for conveying television programs and other digital information to consumers. The flexibility is provided in large part by the packetized transport and multiplex layer of the ATSC system, which includes an important feature that provides program and system information. The ATSC transport layer provides a toolkit of features that can be used to extend the capabilities of the DTV system far beyond what the initial designers might have envisioned. The transport and multiplex layer is based on the MPEG-2 Systems standard [2], which defines system coding at two hierarchical layers: The packetized elementary stream (PES) layer and the systems layer, either in transport stream (TS) or program stream format. The ATSC DTV system uses only the TS format. Section II of this paper provides a tutorial description of the functionality and format of the ATSC transport layer, including the specific constraints on the MPEG-2 Systems standard. Section III of this paper describes a major feature of the ATSC DTV system that extends the MPEG-2 Systems standard to add a Program and System Information Protocol (PSIP). PSIP goes beyond the basic program-specific information (PSI) defined by MPEG-2 to provide a television receiver with additional information necessary for efficient browsing and event selection. The ATSC PSIP standard [3] defines a collection of tables designed to operate within every TS for terrestrial broadcast of DTV. Its purpose is to describe the television channels and scheduled programming events carried in a particular TS. Additionally, information for analog channels as well as digital channels from other TSs may be incorporated. There are two main categories of information in the ATSC PSIP standard, *system information* and *program information*. System information allows navigation and access to the channels within the DTV TS, and program information provides the information necessary for efficient browsing and event selection. Sections IV and V of this paper discuss how the transport layer of the ATSC DTV system provides for private data transport and conditional access (CA) to programs, respectively. Finally, Section VI explains the bitstream interface characteristics of the transport layer of the ATSC DTV system.

Manuscript received June 27, 2005; revised October 10, 2005.

B. J. Lechner, retired, consultant, is at 59 Carson Road, Princeton, NJ 08540-2207 USA (e-mail: tvbernie@ieee.org).

R. Chernock is with Triveni Digital Inc., Princeton Junction, NJ 08550 USA.

M. K. Eyer is with Sony, Bellevue, WA 98004 USA.

A. Goldberg is with Sharp Laboratories of America, Fairfax, VA 22031 USA.

M. S. Goldman is with TANDBERG Television, Bedford, NH 03110 USA.

Digital Object Identifier 10.1109/JPROC.2005.861717

II. TRANSPORT FUNDAMENTALS

The ATSC transport layer is based on the MPEG-2 TS format, as defined by the MPEG-2 Systems standard [2]. The

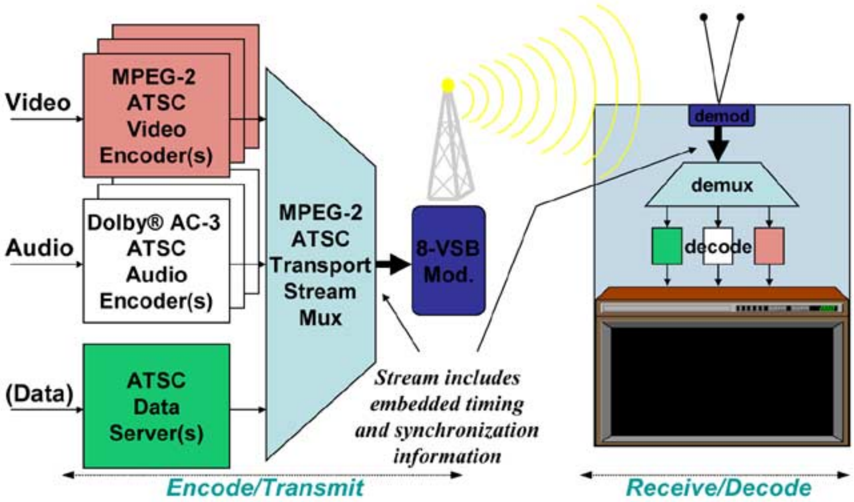

Fig. 1. Example ATSC transmission/reception block diagram showing the MPEG-2 TS multiplexer.

MPEG-2 TS provides a mechanism to encapsulate and multiplex coded video, coded audio, and generic data into a unified bitstream. In order to facilitate parsing of the information contained within the bitstream, in-band control information, known as PSI, is defined. The TS syntax also includes timing information in the form of timestamps in order to enable the real-time reproduction and precise synchronization of video, audio, and data (as applicable). A further design goal of the MPEG-2 Systems standard was to define a TS packetization format that would facilitate real-time transmission and reception of DTV over error-prone physical transmission paths, including over-the-air broadcasting and cable television networks. Fig. 1 contains a block diagram of a generic ATSC transmission/reception system. In typical implementations, the video and audio encoders and the ATSC multiplexer each create output bitstreams in the TS format. The consumer receiver, e.g., an integrated DTV receiver or a digital set-top box, includes inverse TS functions to recover the program information, e.g., a motion picture or a TV program, and deliver it to the decoders, which in turn present it to the viewer.

A. *Multiplex Concepts*

Conceptually, the TS may be represented as a large communications pipe containing one or more smaller pipes. Each smaller pipe represents an MPEG-2 Program. Fig. 2 shows a simple example where the TS carries a single MPEG-2 Program. A Program, in MPEG terms, is an individual program service, such as a single DTV channel. Each MPEG-2 Program comprises one or more program elements, which may include video, one or more audio (e.g., multiple languages), and data streams. Note that a Program may consist of only an audio stream or only a data stream. To avoid confusion with the common usage of the word "program" to refer to a "television program," e.g., a scheduled daily newscast, this paper will refer to television programs as events, and "Program," when used as a noun, will be capitalized to remind the reader that it means an MPEG-2 Program.

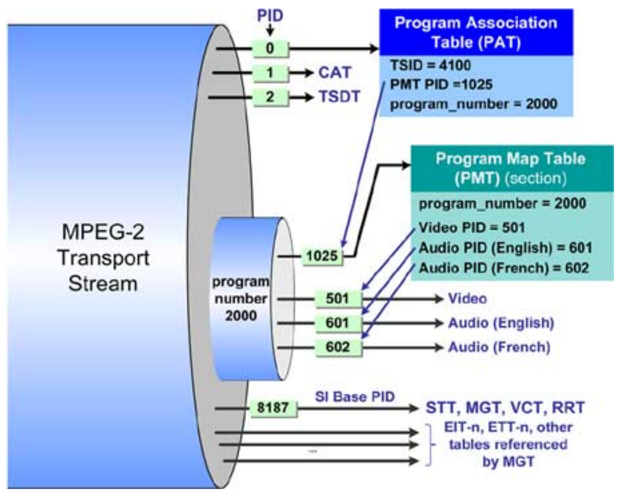

Fig. 2. Illustration of a TS showing multiple layers of multiplexing.

The elementary stream produced by the audio or video encoder is segmented into a series of PES packets (see Fig. 3), typically along frame boundaries in order to facilitate random access to the content. A frame is defined as a progressive picture, both fields of an interlaced picture, or a fixed number of audio samples. An access unit is either the coded representation of a picture or an audio frame. The PES packets, in turn, are further segmented into fixed-length TS packets to facilitate transmission in real time. The process of interleaving the TS packets of more than one Program into a single unified bitstream, while maintaining timing synchronization of each Program contained within, is known as multiplexing. The unified bitstream is called a multi-Program TS (MPTS) and is also referred to as a service multiplex. The MPTS construct enables the deployment of practical, bandwidth-efficient digital broadcasting systems, with each service capable of being delivered at an independent, variable bit rate from other services within the overall fixed bit rate of the MPTS.

The packet identifier (PID), contained in the header of each TS packet (see Section II-B1), is the key to sorting out

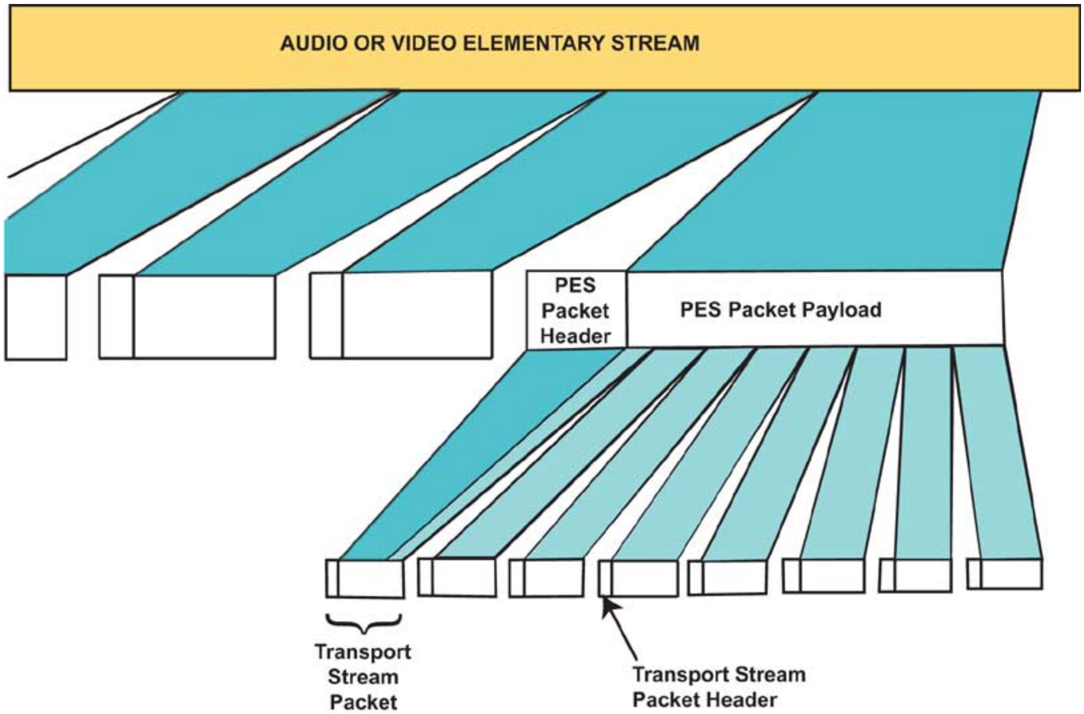

Fig. 3. Packet structure hierarchy of a TS.

the components or elements in the TS. The PID is used to locate the TS packets of a particular component stream within the service multiplex in order to facilitate the reassembly of the payload of each TS packet back into its higher level constructs, i.e., TS packets into PES packets and PES packets into an elementary stream. A series of TS packets containing the same PID includes either a single program element, e.g., a video elementary stream, or descriptive information about one or more program elements, e.g., a PSI table (see Section II-B3).

B. *TS Basics*

The MPEG-2 Systems standard defines the bitstream syntax and the methods necessary for receiving and demultiplexing time-synchronized coded video, coded audio, in-band control information (PSI) that is defined by MPEG standards and other data (including data essence, i.e., data program content) that is not defined by MPEG standards, and that MPEG refers to as private data. The standard includes the definition of packet formats, the synchronization and timing model, the mechanism for identifying content carried in the bitstream, and the buffer models used to enable a receiving device to properly decode and reconstruct the video, audio, and/or data presentation. The information provided by the MPEG-2 Systems standard is sufficient to define the analogous transmission-side functions of synchronizing, packetizing, multiplexing, and transporting DTV services. ATSC standards constrain and extend the MPEG-2 Systems standard to better fit and enhance the practical needs of delivering DTV programming.

An MPEG-2 TS is a continuous series of TS packets (see Fig. 3). A TS packet is 188 bytes in length and always begins with the synchronization byte 0x47. The first 4 bytes (including the synchronization byte) make up the TS packet header. The remaining 184 bytes are available to carry up to 184 bytes of TS packet payload and may include an optional adaptation field.

The definition of the contents of the TS packet payload may differ depending upon the MPEG-2 `stream_type` and the encapsulation method (see Section II-B3).

1) TS Packet Header Fields: In the TS packet header (see Table 1), the PID is a 13-bit value used to identify multiplexed packets within the MPEG-2 TS. Assigning a unique PID value to each component bitstream allows TS packets from up to 8192 (2^{13}) separate bitstreams to be simultaneously carried within the TS. The PID provides a unique bitstream association for each TS packet.

The `payload_unit_start_indicator` is used to signal to the decoder (by being set to "1") that the first byte of either a PES packet or a table section (see Section II-B3a) is located within the payload of the current TS packet. In the case of sections, e.g., PSI and private sections, `payload_unit_start_indicator` field set to "1" also defines the first byte of the TS packet payload to be the `pointer_field`, which indicates the byte offset from the start of the TS packet payload to the beginning of the next PSI or private section. If the `payload_unit_start_indicator` field is set to "0," then the first byte of the TS packet payload is not a `pointer_field`. Instead, the TS packet payload contains the continuation of a previously started PSI or private section along with any necessary stuffing bytes.

Table 1
MPEG-2 Systems TS Packet Header

Syntax	Number of Bits
transport_packet(){	
sync_byte	**8**
transport_error_indicator	**1**
payload_unit_start_indicator	**1**
transport_priority	**1**
PID	**13**
transport_scrambling_control	**2**
adaptation_field_control	**2**
continuity_counter	**4**
if(adaptation_field_control == '10' || adaptation_field_control == '11'){	
adaptation_field()	
}	
if(adaptation_field_control == '01' || adaptation_field_control == '11') {	
for (i = 0; i < N; i++){	
data_byte	**8**
}	
}	
}	

The `transport_scrambling_control` field indicates whether the TS packet payload has been scrambled and, if so, which of two scrambling keys is currently in use. The TS packet header, the optional adaptation field, and the payloads of PSI (see Section II-F), PSIP (see Section III), and null (see Section II-B2) TS packets are never scrambled.

The `adaptation_field_control` field signals the inclusion of the optional adaptation field. The most significant bit of the 2-bit field always indicates the presence of the adaptation field. The least significant bit indicates the presence of payload.

The `continuity_counter` field is a 4-bit rolling counter that is incremented by one for each consecutive TS packet having the same PID (i.e., a separate continuity count is maintained for each component bitstream) to provide a mechanism for recognizing lost packets. There are three special "nonincrementing conditions" defined: 1) when the `adaptation_field_control` field is set to indicate that the TS packet contains an adaptation field only (no payload); 2) when the `adaptation_field_control` field is set to the "reserved" value; and 3) when the TS packet is a duplicate. A duplicate TS packet is defined as the second of two—and only two—sequential TS packets with the same PID (intervening TS packets that have other PIDs are allowed) that are carrying payload and contain identical byte-by-byte contents (except for the Program clock reference (PCR), if present). Duplicate TS packets may be used for additional error resilience purposes. The `continuity_counter` is considered "continuous" if it has incremented by one from the `continuity_counter` value in the previous TS packet having the same PID or when any of the nonincrementing conditions have been met. Otherwise, the `continuity_counter` is considered "discontinuous." Except in the case when the `discontinuity_indicator` flag (defined in the adaptation field syntax) has been set to "1" to signal a discontinuous `continuity_counter`, if a receiver encounters a situation where sequential `continuity_counter` values are discontinuous, then the receiver should assume that some number of TS packets have been lost and take appropriate action.

Two other fields, the `transport_error_indicator` and the `transport_priority`, are also carried in the TS packet header. The `transport_error_indicator` may be used internally by a receiver to indicate to the TS demultiplexer and subsequent processors that at least one uncorrectable bit error exists in the TS packet. The `transport_priority` field, which is typically not used in ATSC TSs, may be used to indicate that a TS packet with the field set to "1" is of higher priority than other TS packets having the same PID.

2) The Null TS Packet: The null TS packet is a special TS packet designed to pad a TS to create a specific overall constant bit rate stream. While individual services within a service multiplex may have variable bit-rate characteristics, the overall TS must have a constant bit rate to operate correctly within practical networks. Null TS packets are transmitted when there are no other TS packets ready to be transmitted (i.e., a continuous stream of TS packets is always maintained). Null TS packets may be added and/or removed by any remultiplexing process within the data path.

Null TS packets are assigned PID value 0x1FFF. The payload of a null TS packet is ignored and may contain any data values, which further aids in the multiplexing add/drop process (e.g., changing the PID value of a TS packet to 0x1FFF without changing its content indicates to downstream devices that they are to treat the TS packet as nonfunctional data stuffing). The `continuity_counter`

Table 2
MPEG-2 Systems Private Section

Syntax	Number of Bits
private_section() {	
table_id	**8**
section_syntax_indicator	**1**
private_indicator	**1**
reserved	**2**
private_section_length	**12**
if (section_syntax_indicator == '0') {	
for (i = 0; i < N; i++) {	
private_data_byte	**8**
}	
}	
else {	
table_id_extension	**16**
reserved	**2**
version_number	**5**
current_next_indicator	**1**
section_number	**8**
last_section_number	**8**
for (i = 0; i < private_section_length-9; i++) {	
private_data_byte	**8**
}	
CRC_32	**32**
}	
}	

of a null TS packet is undefined, carries no information, and should be ignored.

3) TS Packet Payload Data Structures: The TS packet payload field carries the actual data content that is being transported. The MPEG-2 Systems standard defines two fundamental bitstream data structures, the PES packet and the section. The PES packet is used to encapsulate the coded video, coded audio, or coded program data (e.g., a data program of some sort). The term "essence" is often applied to such coded information (video, audio, or data) to distinguish it from data that constitutes descriptive information about the video, audio, or data essence.

The section is used to encapsulate the descriptive information about the coded video, coded audio, or data essence within the service multiplex (e.g., stream type, information needed to extract the streams, program guide information, etc.). A section may also be used to encapsulate private data essence.

a) Tables and Sections: The MPEG-2 Systems standard defines tables that provide "in-band control" information necessary to act on or to further describe the data essence streams within the service multiplex. The logical tables are constructed by using one or more sections. These tables are collectively called PSI. In addition, the MPEG-2 Systems standard provides a mechanism to add additional tables, outside the scope of the standard itself. This structure is known as the `private_section`. ATSC standards, including the PSIP standard [3] (see Section III), use the `private_section` to define compatible extensions to PSI.

The MPEG-2 `private_section` (see Table 2) defines a data encapsulation method used to place private data (that is, data that the MPEG-2 standards do not define, including ATSC-defined sections) into a TS packet with a minimum amount of structure.

A section always begins with an 8-bit `table_id`, which uniquely identifies the table of which the section is part. Another field, the `section_syntax_indicator`, determines whether the short or long form of the `private_section` syntax is used. The short-form section includes a minimal amount of header information and is limited to carrying a payload of at most 4093 bytes in a single section. The long-form section incorporates additional header fields, which allow the segmentation of large data structures into multiple parts or sections. A collection of long-form sections may accommodate $256 * 4084$ bytes of payload (maximum size of 1 045 504 bytes).

The long-form section header contains a `version_number` field, which identifies the revision of the contents of the section. Any time the payload bytes of a section are modified, the `version_number` must be incremented so that a receiver will be able to determine that the contents of the section have changed.

The long-form section contains a `CRC_32` field as the first byte following the last payload byte, which is used for error detection purposes. The receiver's 32-bit CRC decoder (the CRC decoder model is described in MPEG-2 Systems [2, Annex A]) calculates the CRC result over all the bytes that comprise a section beginning with the `table_id` through the last byte of the `CRC_32` field itself. A decoder CRC accumulator result of zero indicates that the section was received without error.

One or more sections may be placed into a TS packet depending on the size of the section. If the section length is smaller than a TS packet payload, then there may be multiple sections contained within a single TS packet. Sections that are larger than a single TS packet are segmented across multiple TS packets. Once the process of packetizing a section commences, a new section will not begin in a TS packet having the same PID until the packetization of the previous section has been completed. When a section does not completely fill a TS packet's payload area and there is no new section ready to begin filling the remainder of the payload area, the remaining bytes of the TS packet are stuffed, or filled, with the value 0xFF. To prevent stuffing byte emulation, the MPEG-2 Systems standard forbids the use of 0xFF as a `table_id` value.

b) PES Packet: The PES packet is used to encapsulate coded video, coded audio, and data elementary streams. Synchronization information (in the form of presentation and decoding timestamps) and other useful information are included in the PES packet header. Elementary streams are each independently carried in separate PES packets; thus, a PES packet contains data from one and only one elementary stream.

The PES packet consists of the PES packet header followed by the PES packet payload. PES packets may have a variable length. A length field allows explicit signaling of the size of the PES packet (up to 65 536 bytes) or, in the case of video elementary streams, the size may be indicated as unbounded by setting the packet length field to zero. When encapsulating data into a PES packet, the elementary stream is first divided into segments, which may be of variable size, and these segments are then encapsulated using the MPEG-2 PES packet syntax. The ATSC DTV standard [1] places two specific constraints on PES packets that encapsulate video elementary streams: 1) a video PES packet shall contain no more than one coded video frame and 2) the packet must be signaled as being unbounded in size by defining the length field as 0x0000.

PES packets carry stream synchronization information in the PES packet header using decoding timestamp (DTS) and presentation timestamp (PTS) fields. The timestamps enable decoding the access units in the correct order and presenting the access units at the correct relative time, i.e., the correct rate, respectively. The DTS and the PTS are each 33 bits long with units of 90-kHz clock periods.

Since the service multiplex is composed of fixed-size TS packets, "stuffing" is needed when there is insufficient data to completely fill the TS packet payload. Stuffing is the process of filling out the remainder of a TS packet with data bytes that carry no useful information, but only take up the remaining available TS packet payload bytes. For TS packets carrying PES packets, stuffing is usually accomplished by defining an adaptation field longer than the sum of the lengths of the data elements in the adaptation field, so that the payload bytes remaining after the adaptation field exactly accommodate the available PES packet data. This extra space in the adaptation field is filled with stuffing bytes. Note that the MPEG-2 Systems standard provides other stuffing mechanisms. For example, to add stuffing bytes within the PES packet itself, the PES packet header length may be extended up to an additional 32 bytes and the header may be stuffed with the corresponding number of stuffing bytes (0xFF).

C. Multiplex Concepts Revisited: An Example

Fig. 4 shows an example of a service multiplex that brings together the concepts described in the previous sections.

The video stream of Program P1 is illustrated to consist of three MPEG-2 TS packets identified by PID 0x1024. Each TS packet has a `continuity_counter` associated with the specific PID that enables a receiver to determine if a loss has occurred. In this example, the `continuity_counter` values begin at 0x3 and end with 0x5 for the video stream of Program P1. The individual TS packets that contain this PID are extracted from the multiplexed bitstream and reassembled, in this case making up part of a PES packet carrying a video elementary stream.

Program P1 also has an associated audio stream of TS packets identified by PID 0x1025. Two TS packets from the audio stream of Program P1 are shown with `continuity_counter` values of 0x2 and 0x3 respectively. Similarly, the TS packet composition of Program P2 is illustrated. In the video stream of Program P2, identified by PID 0x0377, the `continuity_counter` of the second-to-last TS packet is 0xB rather than the expected value of 0x8. This condition indicates an error and the loss of three or more TS packets in the video stream of Program P2. The next expected and received `continuity_counter` value is 0xC as illustrated in the diagram.

In order to present the decoded bitstreams, e.g., video and audio, of a Program in proper synchronism, the decoder needs to recreate the original system time clock (STC) that the encoder used during the coding process. The encoded bitstream contains samples of the original STC, called PCRs, and the decoder uses the arrival time of the PCRs and the PCR values themselves to recreate the STC. See Section II-E1 for more details.

Because of the STC recovery process, TS packets containing a PCR cannot be rearranged in the TS casually. This limitation exists because shifting the relative location of a TS packet carrying the PCR introduces jitter into the data stream, which may cause the STC recovered by the decoder to vary and, in turn, cause problems with video or audio presentation and synchronization.

Also notice the null TS packets (identified by PID 0x1FFF) that were interleaved. These may appear anywhere in the stream and are often used to set the service multiplex to a known, fixed overall bit rate, regardless of the sum of the bit rates of all the Programs contained therein. For illustrative purposes, a value of 0x03 is shown in Fig. 4 for the `continuity_counter` for the null TS packets. In practice any value may be used, as the `continuity_counter` for null TS packets is ignored.

D. Buffer Model (A Basic Introduction to the T-STD)

Merely specifying the TS syntax is not sufficient to create an interoperable system. Receivers are expected to be able to receive, parse, and act on information in real time. A receiver designer must therefore know both the syntax of the data to be received and the timing and buffer-requirement characteristics of the stream.

The MPEG-2 Systems standard specifies these characteristics by describing a "TS System Target Decoder" (T-STD) model. The T-STD is a model of a theoretical receiver that has certain buffering and processing characteristics,

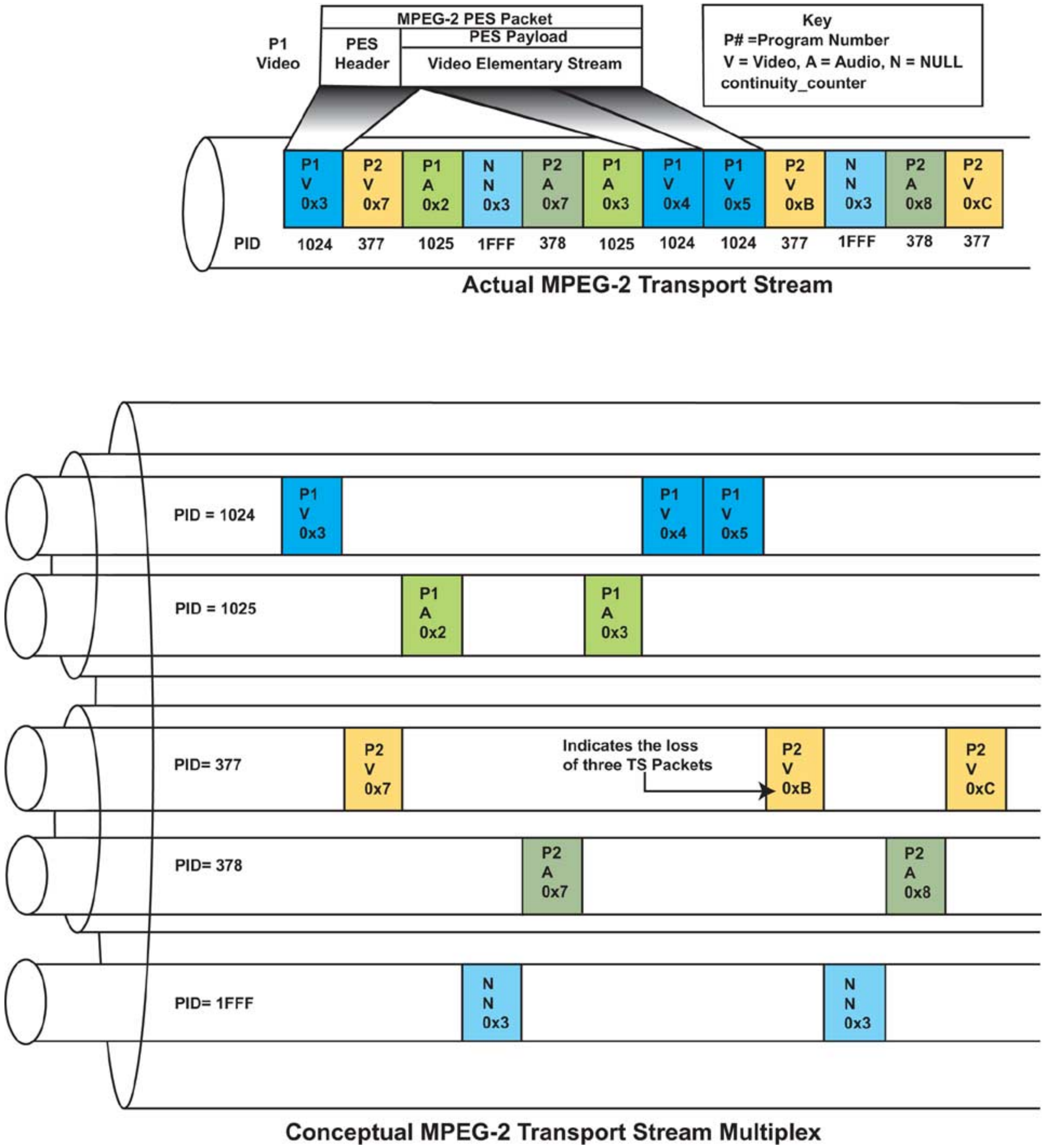

Fig. 4. Conceptual MPEG-2 TS multiplex.

and a TS complies with the MPEG-2 Systems standard only if is receivable by the T-STD without error. Note that the T-STD is actually an encoding constraint on the bitstream.

The T-STD is illustrated in the block diagram shown in Fig. 5.

Fig. 5 shows TS bytes coming in on the left, being demultiplexed (based on PID), and going into various buffers at various rates, before being decoded.

For Video (the top line), data goes into:

1) Transport Buffer (TB_1), at the transmission rate;
2) Multiplexing Buffer (MB_1), at the rate RX_1;
3) Elementary Stream Buffer (EB_1) at the rate Rbx_1.

Then decoder D_1 decodes the video.

For Audio (the middle line), data goes into:

1) Transport Buffer (TB_n), at the transmission rate;
2) Main Buffer (B_n) at the rate RX_n.

Then decoder D_n decodes the audio.

For Systems Data (the bottom line), data follows a similar path to Audio.

The sizes of the buffers, and the rates at which data goes in and out of them, are defined (variously) by the MPEG-2

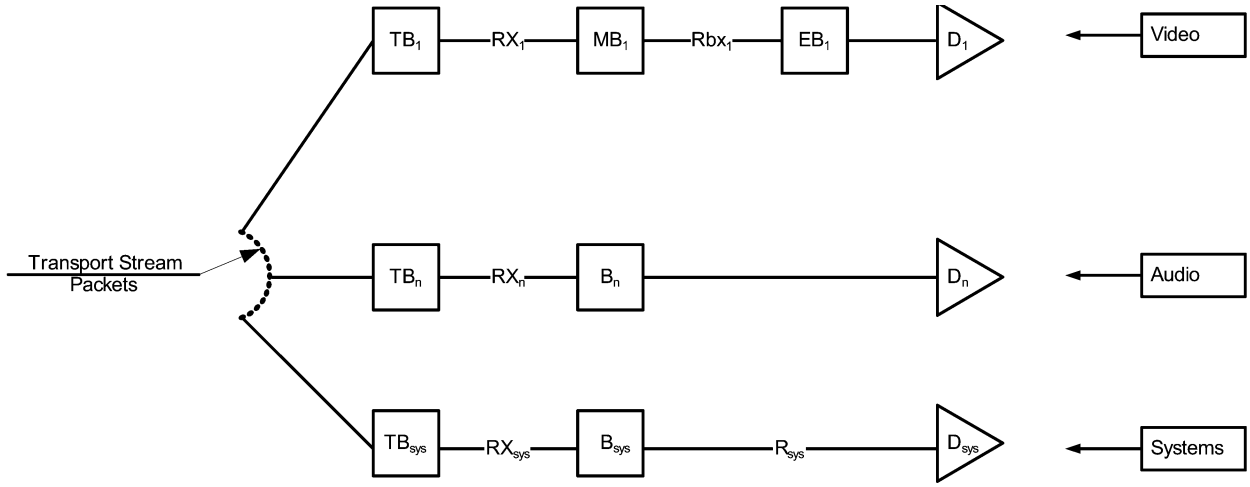

Fig. 5. TS system target decoder notation.

Systems standard, the MPEG-2 video standard [4], and the ATSC digital audio standard [5].

The ATSC DTV system is defined such that TSs neither underflow nor overflow these buffers (except in special cases, e.g., when a Program is being added to or removed from the TS).

The buffer sizes are as follows.

- $\mathrm{TB_1}$, $\mathrm{TB_n}$, and $\mathrm{TB_{sys}}$ have a fixed size of 512 bytes.
- $\mathrm{MB_1}$ has a fixed size of 426 666 bytes for decoding video formats meeting MPEG-2 Main Profile at High Level (MP@HL).
- $\mathrm{EB_1}$ has a fixed size of 1 222 656 bytes for decoding video formats meeting MPEG-2 MP@HL.
- $\mathrm{B_n}$ has a fixed size of 5696 bytes (see [5, Annex A, Sec. 4.4]).
- $\mathrm{B_{sys}}$ has a fixed size of 1536 bytes (see [2, Sec. 2.4.2.3]).

A hypothetical receiver that implements the T-STD shown in Fig. 5 receives TS packets and the TS packets are demultiplexed by PID and reassembled into video, audio, and other streams. Immediately after the demultiplexer, there is a small, fast buffer $(\mathrm{TB_n})$ ready to receive it. Video data then passes into a series of two more buffers ($\mathrm{MB_1}$ and $\mathrm{EB_1}$), each larger and slower than the previous. Other data passes into one bigger, but slower, buffer ($\mathrm{B_n}$).

After passing through the two (three for video) buffers, the final buffer is defined to have an unbounded output rate. That is, similar to audio and video decoder modeling, the T-STD is modeled having an instantaneous decode. Receiver designers must take note of this and design hardware with larger and faster buffers than described by MPEG.

1) Typical Operating Conditions: Consider video buffer $\mathrm{EB_1}$, for example. The MPEG-2 Systems standard requires that this buffer be filled at a rate no faster than 105% of the video elementary stream rate, and drained one picture (field or frame) at a time instantaneously at the picture decode time signaled by the DTS. This will occur typically at the transmitted television frame or field rate. However, note that the display field or frame rate may be different, e.g., film-based material transmitted at 24 frames per second that is displayed at 30-interlaced or 60-progressive frames per second.

In accordance with the MPEG-2 video standard, pictures are coded in three types: intraframe (I) pictures; predicted (P) pictures; and bidirectionally predicted (B) pictures. I pictures are based on information only from the current picture; P pictures are based on information from the current picture and information from a previous I or P picture; and B pictures are based on information from the current picture and information from both a previous I or P picture and a future I or P picture. In order to decode B pictures, the future I or P picture information must be available to the decoder. For this reason, the pictures are transmitted out of order. Current practice is to transmit the pictures in the order in which they are to be decoded.

Fig. 6 shows that, just before the first I picture is decoded, the buffer is approaching full. Immediately after the I picture is decoded, the buffer is nearly empty. As time passes, B pictures are decoded, but since they take fewer bits than I pictures, the fullness of the buffer increases. At the time the first P picture is decoded, the buffer is also nearly full. Immediately after the P picture is decoded, the buffer is emptied, but not as much as after the I picture.

This pattern continues. Note that encoders are designed to make the most out of the buffering available—complicated images and I and P pictures require more bits than simpler images and B pictures.

Also note that the buffer may not fill at a constant rate as shown in Fig. 6 when variable rate PES streams are included in the TS.

2) Transport Buffers (TB_n)*:* In contrast to the $\mathrm{B_n}$'s discussed above, $\mathrm{TB_n}$'s are much smaller and much faster (on input), and feed into the larger, slower buffers. The size of $\mathrm{TB_n}$ is fixed at 512 bytes for all streams (see [2, Sec. 2.4.2.3]).

The small fast buffer was included in the T-STD buffer model in order to deal with demultiplexer implementation issues relating to memory bandwidth and data transfer rates. Since the 512-byte size for $\mathrm{TB_n}$ is less than three TS packets, there cannot be more than two TS packets in a row for an audio stream. This is because of the drain rate $\mathrm{RX_n}$, which is limited by the size of the main audio buffer $\mathrm{B_n}$. Multiplexers work very hard to assure that this constraint is met.

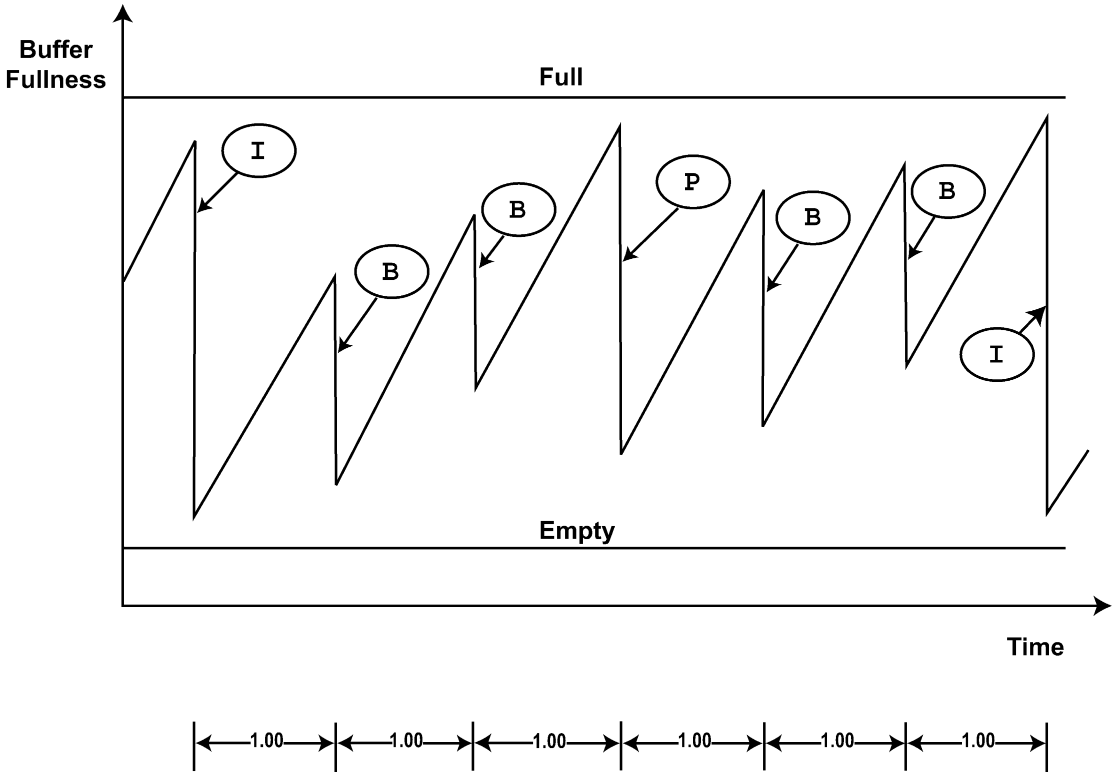

Fig. 6. Video buffer occupancy over time.

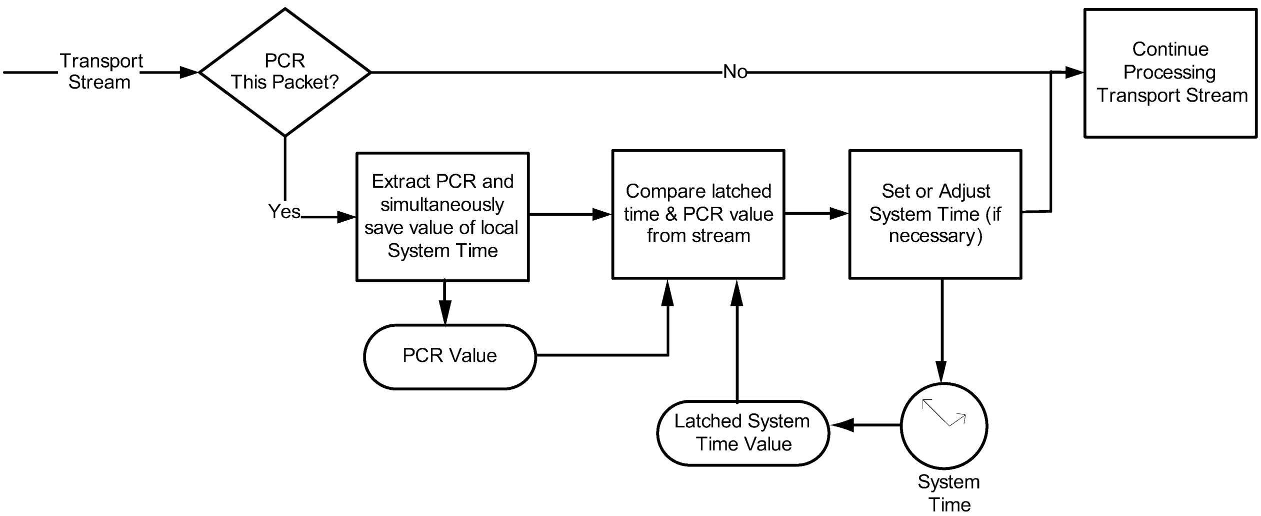

Fig. 7. STC recovery processing (schematic).

E. Synchronization and Timing

The MPEG-2 Systems standard describes a method for signaling synchronization between access units in multiple elementary streams. Practically, for audio/video ("television") programming, this generally means audio-visual synchronization. The encoder attaches a PTS to video pictures and to audio frames (audio presentation units). The PTS is a 33-bit value in cycles or "ticks" of a 90-kHz STC. Fundamentally, the PTS tells the receiver what time to display a picture or output an audio frame. In order to present audio and video each at the correct time, the receiver needs to have its STC set to the correct time.

1) Recovery of the STC: The STC provides a reference for all decoding and display operations in the receiver. A receiver must have enough information to set its clock. When the TS is created, periodically (not less often than once each 100 ms) a PCR is included in the TS packet adaptation field for each Program. This is the value of the STC at the encoder. The receiver adjusts its clock based on the PCR values. Fig. 7 illustrates the process whereby the receiver recovers the PCR from the TS and uses it to set or adjust its local STC.

As shown in Fig. 7, the receiver examines each TS packet as it arrives to determine whether it carries a PCR field for this Program. If a TS packet contains a PCR, the receiver does two things immediately: 1) stores (latches) the current STC in a register and 2) stores (latches) the PCR value from the TS packet in a separate register. As soon as practical, the receiver examines the difference between the latched PCR from the TS packet and the latched local STC value. If they are identical, the local STC is running at the correct speed (it exactly matched the encoder's STC). If they are slightly different, the receiver adjusts the speed of the local STC (slightly speeds it up, if the STC value was less than the PCR value, or slightly slows it down, if the STC value was more than the PCR value).

However, if the difference between the times is large (as in the case of a discontinuity—perhaps a splice, or a channel change, or reconfiguration of the encoder), then the receiver disables its audio–video synchronization processing and sets ("jam loads") the local STC value with the value of the PCR in the next TS packet that carries a PCR for that Program. After the STC has been reset in this manner, and after sufficient time for the data buffered before the STC reset to be decoded and displayed, the receiver reenables its audio–video synchronization processing.

By this process, the receiver "recovers," or creates, an internal copy of the encoder's STC. This is vital to the audio–video synchronization processing described below. Additionally, the STC hardware in the receiver is typically used to derive audio and video timing clocks necessary for correct audio pitch and video color display.

2) Synchronization: In systems using MPEG-2 TSs, audio and video (and other data streams, where applicable) are not inherently synchronized to each other when placed into the TS. However, each component, e.g., picture or audio frame, of each MPEG-2 Program carried by a TS contains a PTS. The receiver uses these PTSs along with its STC to present the audio and video at the correct time (with respect to the STC). That is, instead of direct "audio-to-video" synchronization, there is "audio-to-system time" and "video-to-system time" (and "data-to-system time") synchronization.

The MPEG-2 Systems standard allows a PTS to be carried in each PES packet header. The PTS indicates, relative to the STC, when the first access unit carried in the PES packet payload should be presented. The ATSC DTV standard requires that each video PES packet contain exactly one picture, and that each PES packet header contain a PTS. Thus, each picture has an explicit presentation time.

The receiver then presents each picture (and each audio frame) as close to its PTS-signaled presentation time as possible. For video in particular, this may not be precisely when the STC equals the PTS value because the beginning of a video scan period in a receiver may not correspond precisely to the beginning of the ideal presentation time.

According to the MPEG-2 Systems standard, DTSs are optional and most, if not all, receiver implementers consider them superfluous. A receiver can assume the timing of the DTS based on the need to satisfy the corresponding PTS for the same access unit. This is easily done since current coding practice is to transmit video access units in the order in which they are to be decoded. In a similar fashion a receiver can also assume the PTS time for a frame that does not explicitly contain a PTS by knowing the frame rate, the PCR (ignoring PCR discontinuities) and the time of the previous PTS.

F. PSI

The MPEG-2 Systems standard defines tables that provide "in-band control" information necessary to act on or to further describe the video, audio, and data essence streams within the TS. The logical tables are constructed by using one or more sections. These tables are collectively called PSI. PSI provides data necessary to locate specific Programs or program elements in the service multiplex. Five PSI tables are defined: the Program association table (PAT); the Program map table (PMT); the CA table (CAT); the network information table (NIT); and the TS description table (TSDT). Fixed PID values are assigned to some PSI tables, including the PAT, but not the PMT. The PAT is the "base" table for PSI. If it did not use a fixed, known PID, "tuning" (parsing the stream) would take longer and channel acquisition would be slower. The reason the PMT does not have a fixed PID is that the tuning is hierarchical. A single PAT may point to many Program map sections and the number of program map sections depends on specific usage.

The PAT provides a complete list of all of the MPEG-2 Programs (services) within the MPTS or service multiplex. The PAT establishes a relationship between each MPEG-2 Program, via the `program_number`, and its corresponding program map section (properly defined as `TS_program_map_section`), via the PID value assigned to the corresponding program map section. TS packets that contain the PAT are assigned to PID 0x0000. Fig. 2 illustrates the PAT and a PMT in a TS and diagrams the relationship between them.

Each program map section contains the mapping between an MPEG-2 Program and the program elements that define the Program (this mapping is called a program definition). Specifically, a program definition establishes a relationship between an MPEG-2 program number (a logical number used to uniquely identify the Program) and the list of the PIDs that correspond to the individual program elements comprising the Program. The PMT is defined as the complete collection of individual program definitions within the TS, with one `TS_program_map_section` per Program. The PMT is unique among the PSI tables in that parts of its contents may be carried in different component bitstreams (i.e., within TS packets that have different PIDs). This simplifies the addition, deletion, or modification of the PSI for individual Programs, as each can be altered independently. This also simplifies the demultiplexing process since only the relevant portions of the TS need to be parsed by the receiver. Each of the other PSI tables is required to be contained in its own unique bitstream (within TS packets having a single, unique PID).

Whenever component bitstreams of a Program are scrambled (i.e., the contents are only decodable with the use of a CA system process), a CAT must be present in the TS. The CAT associates aspects of the CA system, such as access rights sent in entitlement management messages (EMMs), with the scrambled streams. TS packets that contain the CAT are assigned PID 0x0001. CA systems provide scrambling of Programs or individual program elements along with end user authorization. While Programs or individual program elements may be scrambled, the PSI tables are never scrambled. The MPEG-2 Systems standard does not define the contents of the CAT payload. For details of how the ATSC defines its CA system; see Section V.

The function of the NIT is to carry information that applies network-wide (i.e., to all service multiplexes in the delivery/emission network). The ATSC DTV standard does not specify the use of an NIT.

The function of the TSDT is to carry descriptors that apply to an entire service multiplex. The ATSC DTV standard neither constrains nor specifies the use of the TSDT.

III. PSIP

As discussed in Section II, audio and video programming is transported and labeled using the TS PSI defined by the MPEG-2 Systems standard [2], as modified by the ATSC DTV standard [1]. Although the PSI defined in the MPEG-2 Systems standard provides information about the contents of a TS, the developers of all world DTV standards, including the ATSC DTV standard, recognized that the mechanisms defined by MPEG-2 were not by themselves sufficient to provide a user-friendly environment (for both broadcasters and viewers) for a broadcast television service (MPEG-2 TSs are used for other purposes). In this section, we describe the features and functions of the ATSC PSIP standard [3], which standardizes the transmission of additional navigation information as well as data usable to create an electronic program guide (EPG). Simply put, PSIP defines the manner in which the broadcaster lets the DTV receiver know such things as the name of the channel and the name and description of current and future programming events on that channel. PSIP is actually much more than that.

PSIP defines "system information" (sometimes called "service information" or just SI) for the ATSC DTV system. The ATSC PSIP standard describes a method for delivery of program guide and system data tables carried in any compliant MPEG-2 TS. The primary purpose of PSIP is to facilitate acquisition of and navigation among the analog and digital services available to a particular receiver or set-top box, but it also provides a platform for the data structures needed to support other applications such as data broadcasting.

Since the ATSC DTV system is designed to support terrestrial over-the-air broadcast and since there is no guarantee that any particular broadcast signal will be available at any particular receiving location, it is necessary that each broadcaster transmit the PSIP information required to identify his own TS and its contents. Receivers will then collect the PSIP information from each of the broadcast TSs that it can receive and build a database of information that can be used to implement the acquisition and navigation functions. By FCC regulation, PSIP data is required for digital terrestrial broadcasts in the United States, and cable operators are required to support it as well for the benefit of cable-ready DTV receivers.

A. *The Naming, Numbering, and Navigation (N^3) Problem*

PSIP was developed to address three primary issues related to DTV, none of which was adequately addressed in the MPEG-2 Systems standard.

- Channel names. Broadcasters wanted to label each of their digital channels with a user-friendly name, such as the call letters of the station. Channel names also benefit the user by providing another way to find the channel of interest.
- Channel numbers. Broadcasters realized they wanted the flexibility to be able to assign channel numbers independently of the actual RF channel used to broadcast the digital signal. Such flexibility would allow any digital service to be associated with the same RF channel number that the broadcaster has used for years for his analog NTSC service. For the television-viewing public, such a numbering scheme would group the analog and digital channels of a given broadcaster together on the TV dial.

  The fact that the digital signal carried in a single 6-MHz slice of RF spectrum could deliver more than one "television channel" created a second number-related problem: how should the "subchannels" be numbered and referenced in electronic and printed program guides?
- Program guides. Broadcasters wished to be able to deliver data to support EPGs so a digital receiver would be able to display the current and future television program schedule.

These requirements came to be known as the "naming, numbering, and navigating" or N^3 problem. The term "navigation" in this context refers to the need to support a user's need to "navigate" or find his or her way around among the many digital service offerings. Part of navigation involves being able to go directly to a desired programming service given a well-known or advertised reference. Another part of navigation involves supporting the familiar "channel grazing" or "channel surfing" activity users often enjoy.

1) Defining the (N^3) Requirements: The PSIP standard was developed by the ATSC T3/S8 Transport Specialists group in 1997. Prior to assembling the new specification, a comprehensive set of design requirements was drafted. The members of T3/S8 were aware of the work being done on service information in Europe by the Digital Video Broadcasting (DVB) consortium, and in fact some T3/S8 members participated in that effort. T3/S8 met with broadcasters, cable television operators, and television receiver manufacturers to document the following system requirements to address the N^3 problem.

- The system had to support direct access to any channel, meaning that the navigational model needed to support the ability to access any analog or digital channel by direct entry of its channel number. Such an access method was deemed a necessary user feature. It is required to support devices that use remote control commands to tune a receiver in order to make an unattended recording. This design decision precluded an approach where navigation of a GUI (moving a cursor on-screen and selecting a channel from a list of choices) was the only method offered.

- The approach had to support grouping of selected digital services with an existing analog service, or with digital services on other multiplexes. For a period, broadcasters will operate an analog channel in addition to a digital multiplex. The navigation model had to include a grouping concept to support channel surfing within a set of related analog and digital channels.
- The approach needed to support the preservation of channel branding. The system had to allow a broadcaster, when starting a new digital service, to associate the new programming with the channel label that had been used to establish identity in past years of advertising.
- The resulting standard had to be harmonized with cable television standards. There was a clear recognition that cable set-top boxes and cable-ready DTV receivers would also employ navigational and channel naming methods. Typically, cable headends carry local terrestrial broadcast channels. PSIP designers had to consider the needs of cable operators to make sure harmony with methods defined for terrestrial broadcast was maintained. A big hurdle early on in gaining support for the new ATSC PSIP standard was the fact that cable operators had already installed digital set-top boxes (which they owned and leased to subscribers) that employed multiple different proprietary navigation systems.
- The approach had to recognize and accommodate the flexibility of digital transport. In a terrestrial broadcast scenario, just as with a cable or satellite system, the use and contents of any given TS varies from time to time. The system design had to gracefully accommodate such changes.
- The implementation solution had to support printed program guides. Even though electronic means are now common, printed program guides still exist. In addition, even though the "look and feel" for the receiver is not specified, whatever channel numbering and labeling method is used in the receiver must align with channel numbering and labeling found in printed guides. The printed guides must be able to clearly describe cable channels as well as programming received via terrestrial broadcast.
- The design needed to allow a broadcaster to "package" or "market" some services separately from others on the multiplex. For example, as a public service, the holder of a digital broadcasting license may offer a couple of spare megabits of standard-definition (SD) TV bandwidth to a college or community access channel, or to a government affairs (city politics) channel. It had to be possible for that operator to label that channel separately from the other services offered on that multiplex.
- The design had to support channel naming, which involves downloading to the receiver the textual name or call-sign associated with each program source. Processing in the receiver of this text would be optional.
- The approved approach could not preclude the development or adoption of new navigation paradigms. The approach must be flexible and extensible enough so that alternative navigation methods may be accommodated in the future. For example, the solution must also accommodate the presence of interactive services.

The PSIP standard also could not preclude or limit the development of data broadcasting services. While the initial design did not need to explicitly support data services, it had to support backward-compatible extensions for data. In the years since PSIP was completed, ATSC has developed and adopted a series of standards for data broadcasting. See "Data Broadcasting and Interactive Television" [6] in this issue.

a) Terrestrial Broadcast Requirements: Certain requirements were specific to the terrestrial broadcast application.

- The system had to accommodate terrestrial broadcast translators. Broadcast translators are used in some communities where terrain such as a hill or mountain prevents the transmitter's main signal from reaching the full service area. In such cases, the MPEG-2 TS may be transmitted on more than one broadcast carrier frequency. The receiver must recognize multiple instances of one TS as alternate physical access points to the same services. The system needed to support the capability of a receiver to provide consistency of labeling (channel name and number) for each instance.
- An additional requirement related to translators was that it must not be necessary to alter the TS in any way prior to retransmission at the translated frequency. In other words, the PSIP data in the untranslated signal must be usable as is in the translated one.
- The system needed to take into account movable terrestrial broadcast receiving antennas. A receiver could encounter one digital broadcast on frequency X at one point in time and a different TS on the same frequency X at another time due to propagation conditions or the reorientation of a movable antenna. Designers were also aware that some people have TV receivers in their recreational vehicles for use while driving or camping.

b) Cable Requirements: Other requirements were specific to the cable television application.

- The cable system operator had to be able to label digital services independently of the RF channel number used to transmit them on cable. This requirement came from the desire by the cable system operators to be able to change the specific RF carrier frequency used to deliver any particular service without affecting the user's notion of the channel number.
- It had to be practical for the cable system operator to create TSs by assembling services from various input streams received via any transmission medium. Because effective data rates on cable can be nearly double those of terrestrial broadcast, services from two terrestrial multiplexes may be combined into a single cable TS. The approach had to allow the cable operator to perform such remultiplexing with a minimum of difficulty.

- It must be possible to label a channel with a three-digit channel number, for consistency with the existing common practice in digital cable systems of numbering channels within the range of 1 to 999.

c) Desirable Features: Some features of the system were not strictly necessary but conformance was clearly desirable.

- The approach should incorporate familiar paradigms for operation. The new services should fit easily within the mental model already used by the TV-watching public. This means that channel surfing and the familiar numbering of channels should work pretty much as they do for analog television receivers.
- The naming method should support naming of analog channels as well as digital. This would allow the most integrated look and feel for the receiver during the transition to full-digital transmission.
- The approach should use standards wherever possible.
- Finally, the approach should harmonize with emerging satellite and multichannel multipoint distribution service (MMDS) standards. The chosen solution for broadcast should also recognize other transmission media.

2) Solving the N^3 Problem: The ATSC PSIP standard solves the N^3 problem. It introduces two important features: 1) the two-part channel number; and 2) the concept of "virtual channels." PSIP uses the MPEG-2 Systems private section syntax (see Section II-B3b) to define a collection of tables and descriptors that are used to deliver both the required and optional PSIP data to the receiver.

a) The Two-Part Channel Number: In analog broadcast or cable television, if a user selects channel 4, a receiver knows to tune to the frequency associated with RF channel 4—the 66–72-MHz band. The correlation between the user's notion of a channel number and the 6-MHz frequency band carrying the RF modulated signal has long been established by well-known standards. For terrestrial broadcast, the FCC rules [7] (see 47 CFR 73.603) define the mapping. For cable, CEA-542-A [8] applies.

The situation has changed, however, with the advent of digital transmission. PSIP offers digital broadcasters the freedom to define channel numbers independently of the RF frequency used to carry the signal. The PSIP architecture is designed around the concept of "virtual channels." A virtual channel is called "virtual" because its definition is given by indirect reference through a table section called a virtual channel table (VCT). So when a viewer selects channel 4, he or she is actually selecting the portion of the VCT associated with user channel number 4. The definition of the channel as given in the VCT includes its textual name, channel type (analog audio/video, digital audio/video, audio only, data), and the channel number the user may use to access it.

The PSIP protocol introduces a new navigational concept, the "two-part" channel number. Broadcasters declared the need, as new digital services are introduced, to retain the brand identity they currently have as a result of years of marketing and advertising. For broadcasters in the United States, the first part of the two-part number (called the "major" channel number) is required to be the same as the CEA/FCC RF channel number already in use for the analog NTSC service. The second part of the number (called the "minor" channel number) identifies one service within the group of services defined by the major number. From the point of view of the user, where before there was just channel 4, now there may also be channels 4.1, 4.2, 4.3, and so on.

Note that the delimiter, or punctuation, between the major and minor numbers is not specified in any current standard. Common practice seems to be gravitating toward use of the decimal point as the preferred delimiter. Since the delimiter is not part of the PSIP data, broadcasters have no choice except in video overlays, printed guides, and Web sites, where some broadcasters use the hyphen and others use the decimal point.

b) Virtual Channels: Virtual channels are the cornerstone of the ATSC solution to the N^3 problem for DTV. When a viewer tunes a terrestrial broadcast or cable receiver to a digital channel, the channel number he or she uses to identify that channel is not tied to the RF carrier frequency carrying that channel. Instead, the user's notion of the channel number is a parameter set by the broadcaster or cable operator in PSIP data.

c) What's in a Virtual Channel?: To see what the VCT adds, examine for a moment what the world would look like without it, or without any service information aside from that defined in the MPEG-2 Systems standard. Consider a receiver built to only use the MPEG-2-defined PSI tables: the PAT and the PMT. Assume this receiver could find a digital multiplex on a particular RF channel on its own because it knows the frequency plans and modulation methods in use in a particular region of interest. Given that a particular MPEG-2 TS is found at a certain carrier frequency, the receiver knows the following information from the PSI tables:

- the TS's ID (TSID);
- that the TS includes some number of MPEG-2 Programs;
- the MPEG-2 "program number" associated with each Program;
- by inspection of the PMTs, the composition of each Program in terms of video, audio, data, and metadata streams.

What can the receiver display to the user to help in navigation? First of all, the RF channel number of the carrier on which the TS was found may be helpful. The receiver could also somehow use the MPEG-2 program number associated with each service, but that number may not be user-friendly—it can range from 1 to 65 535. Furthermore, it could change over time and it may be duplicated by other broadcasters in the same region. A receiver could number services in their order of appearance in the PAT: the first one could be called Program #1, the second #2, etc.

In contrast, the PSIP VCT adds the following information to each service definition:

- a one- or two-part channel number that can be used for channel surfing or for reference to printed program guides;
- the name of the channel (up to 7 characters), which typically indicates a broadcaster's call letters or for

cable channels, the name of the service (for example "KNSD," "HBO," "C-SPAN," etc.);
- the type of channel, whether it is analog NTSC television, a DTV service, an audio-only service, or a data-only service.

The VCT gives additional information the receiver can use to help the user find and/or identify the service or to properly display data relating to it. This information includes:
- the channel's source ID, a 16-bit number that links this channel with the EPG data;
- a flag indicating whether or not the channel is "hidden," meaning it is only available via an application or only to equipment owned by the broadcaster or cable operator;
- a flag indicating whether or not the channel should be included in EPG displays (note that some channels may be "inactive" meaning that they are currently unavailable (hidden), yet they may appear in EPG displays);
- a flag indicating whether the channel may require a subscription for viewing;
- an optional text string that also provides the name of the channel without the seven-character length limit;
- zero or more descriptors providing further information or attributes.

The VCT provided with terrestrial broadcast TSs also includes a list of the PID values associated with the audio and video elementary streams. This can help decrease the time it takes for a receiver to present audio and video when the channel is first tuned.

B. *Main PSIP Tables*

In this section we describe the core set of SI tables specified in the ATSC PSIP standard: the system time table (STT); the master guide table (MGT); the cable VCT (CVCT) and terrestrial VCT (TVCT); the rating region table (RRT); the event information table (EIT); and the extended text table (ETT). The directed channel change tables (DCCTs) will be described in Section III-F.

All of the table sections defined in the ATSC PSIP standard are derived from the "long form" syntax of the MPEG-2 `private_section` syntax. MPEG uses the term "private" in this context to mean "defined outside MPEG." Each of the PSIP tables includes one additional 8-bit field after the MPEG-defined header bytes, the `protocol_version` field. The `protocol_version` offers yet another extension for the `table_ID` in that it allows, in the future, ATSC to define a type of table section entirely independent of an existing table section definition while it shares the same `table_ID` value. For the currently defined tables, `protocol_version` is zero.

1) STT: The STT is a small section that is sent once per second to set the receiver time base referenced in the other tables, particularly the EIT, and to provide the current time of day. It also indicates whether or not Daylight Saving Time is in effect and signals the day and hour for transitions into or out of Daylight Saving Time. Time in the STT is represented as the count of seconds that have occurred since 00:00:00, 6 January 1980. This date marks the "beginning of time" from the perspective of the time base used by the global positioning satellite (GPS) system. To convert this count into time of day, one must account for the number of leap seconds that have occurred since the start of GPS time and the present. The STT provides that number in a parameter called "GPS-to-UTC offset."

The ATSC PSIP standard stipulates that the count of GPS seconds delivered in the STT must be correct to within plus or minus one second. The leap seconds count must be exact. Achieving this level of accuracy is straightforward, but there are several sources of potential error within the system.
- Accuracy of the reference clock used to feed the STT generator. If a GPS receiver is used, accuracy to 1 μs is possible, since such GPS receivers often produce output pulses to that level of precision. If other clock sources are used, for example, a PC's internal clock, accuracy could be much worse and could drift over time.
- Latency between the clock reference and the actual time at which the STT is transmitted by the broadcaster. This will include delays in the STT generator, the PSIP generator, the various multiplexers that assemble the TS, the studio-to-transmitter link, and even the exciter in the broadcast transmitter.

Very good precision can be achieved by precompensating for these delays.

2) MGT: The PSIP MGT is called "master" because it refers to all other PSIP tables (except the STT). The purpose of the MGT is to provide a convenient reference for receiving devices to the various instances, i.e., occurrences, of PSIP tables present in the TS. The MGT offers the following information for the benefit of receivers.
- A list of all PSIP tables present in the TS. Tables other than those defined in the PSIP standard may be listed as well. For example, the ATSC data broadcast standard [9] uses the MGT to announce the presence of the data event table (DET) and the ETT associated with it, and the long-term service table (LTST).
- For each type of table listed, the PID value for the TS packets that are used to carry it.
- The version number of each type of table. In some cases, one MGT table type represents more than one table section or even more than one instance. For that case, all tables and table sections must share the common version number indicated in the MGT.
- The total number of bytes transmitted in all transmitted sections of tables of each type. The size parameter is helpful to the receiver so that it can determine beforehand whether or not necessary storage resources are available, and if available, how much storage must be allocated.

3) VCT: One of the fundamental pieces of data needed by receivers is a list of available services, including information about each, such as the name of the service and the channel number it is identified with in printed guides and in electronic guides distributed over other media, e.g., the Internet. Such a list is an essential part of the construction of an EPG. The VCT in PSIP fulfills this function. Knowledge of the VCT is essential even for a receiver not supporting an EPG feature since it is the VCT that enables all receivers to identify digital services in a consistent and user-friendly manner.

The ATSC PSIP standard defines two separate types of VCT, one specifically for terrestrial broadcasting and the other for cable. Having two separate types of VCTs allows system operators to define different channel numbers when navigating on cable as opposed to terrestrial broadcast, a practice common in analog distribution. The VCT for terrestrial broadcast is the TVCT and the one for cable is the CVCT. Beyond the header bytes common to all of the PSIP table sections, the structure consists of a count of the number of virtual channels to be defined in the section, followed by each definition itself. Following the channel definitions is a field indicating the length of any descriptors that might be included in the section, followed by the descriptors themselves (see Section III-C for a discussion of descriptors). Finally, the 32-bit CRC terminates the section.

The heart of the VCT structure is the definition of each virtual channel. Parameters comprising each virtual channel definition include:

- the "short" name of the virtual channel, an uncompressed text string of up to seven characters;
- the major and minor channel numbers of the virtual channel;
- an indication of the modulation mode used by the carrier delivering the virtual channel;
- the TSID (or, for analog channels, the transmission signal ID) associated with the virtual channel;
- for digital services, the MPEG-2 program number associated with the virtual channel;
- an indication of the existence and location of an ETT giving further textual description of the virtual channel;
- several flags providing attributes of the virtual channel, such as whether or not it is visible to consumer receivers and whether or not program guide data for the channel should be shown;
- an indication of the type of service associated with the virtual channel, such as analog or DTV service, or audio-only or data-only service;
- the source ID value associated with the virtual channel;
- zero or more descriptors pertinent to the virtual channel.

Fig. 8 illustrates how the VCT references digital programming services in the same TS as that carrying the VCT itself, and shows that it can reference other digital TSs and analog broadcast signals as well. In the example, a TS with TSID value 0x10FD is shown in the upper left. This TS carries the VCT outlined in the lower left. The first two Programs listed in the VCT reside within this same TS, while the second pair are carried in a TS identified with TSID value 0x10FF. The fifth virtual channel in the VCT is an analog television service. In this case, the TSID is the transmission signal ID normally found in the XDS packet on line 21 in the vertical blanking interval (VBI) of the NTSC signal. The transmission signal ID fulfills the same function in linking an analog signal to the VCT as the TSID does in linking a digital stream to the VCT.

4) RRT: The RRT defines the rating parameters, called rating "dimensions" and "levels," that are used to convey program content advisory information (commonly called V-chip data). A TS can carry as many RRTs as there are rating regions in use. See the description of the ATSC content advisory system in Section III-E for a complete discussion of the RRT.

5) EIT: EITs provide event (i.e., television program) description and schedule information for any analog or digital channel listed in the VCT. EIT information consists of the start time and duration of each event, its title, linkage to a textual description of the event, and an optional list of descriptors that can give further information pertinent to the event such as its content advisory and the caption services available.

EIT sections use the standard long-form MPEG-2 `private_section` syntax. In the table header the `source_id` appears in the position of the `table_id_extension`. This allows multiple distinct EIT instances, each corresponding to a different virtual channel (linked by `source_id`), to appear in TS packets with common PID values. The `source_id` is what allows receivers to see the instances as separate from one another, and, by matching `source_id`s found in the table headers, to reassociate different parts of sectioned EITs.

Each EIT instance, then, gives a portion of the program schedule for the programming service identified by `source_id`. The message body of the EIT consists of two "for" loops. The first contains event records and the second is composed of the additional descriptors loop found in most PSIP tables. The number of event records is given by the parameter just in front of the start of the loop, `num_events_in_section`.

Any given EIT describes a 3-hour block of programming, where the 3-hour blocks are aligned to absolute UTC—0, 3, 6, 9, 12, 15, 18, and 21 hours. Events that run for any period within the 3-hour time slot must be included. This means that events starting, ending, or continuing throughout the 3-hour period are included. When an event is included in more than one EIT (EITs for different 3-hour slots), the same `event_id` must be used in each.

An event record consists of information pertaining to an event. Information given about each event includes:

- its starting time and duration;
- title;
- flags indicating whether or not descriptive text exists, and if so, whether it is located in this same TS or in another TS carrying the actual event;
- the `event_id` field used to uniquely identify this event on the virtual channel and also link the event to the textual descriptions carried in the ETTs;
- any descriptors pertinent to the event, such as content advisory descriptors or caption service descriptors.

6) ETT: PSIP defines a method to include multilingual text in the digital TS. The bulk of this text is most often event descriptions, giving a few words, a sentence, or a few sentences of information about the contents of a given event. Depending on the type of event, the text can include the title of an episode, a synopsis of the story line, the names of actors, the year of production, or anything the provider of the EPG data wishes to include.

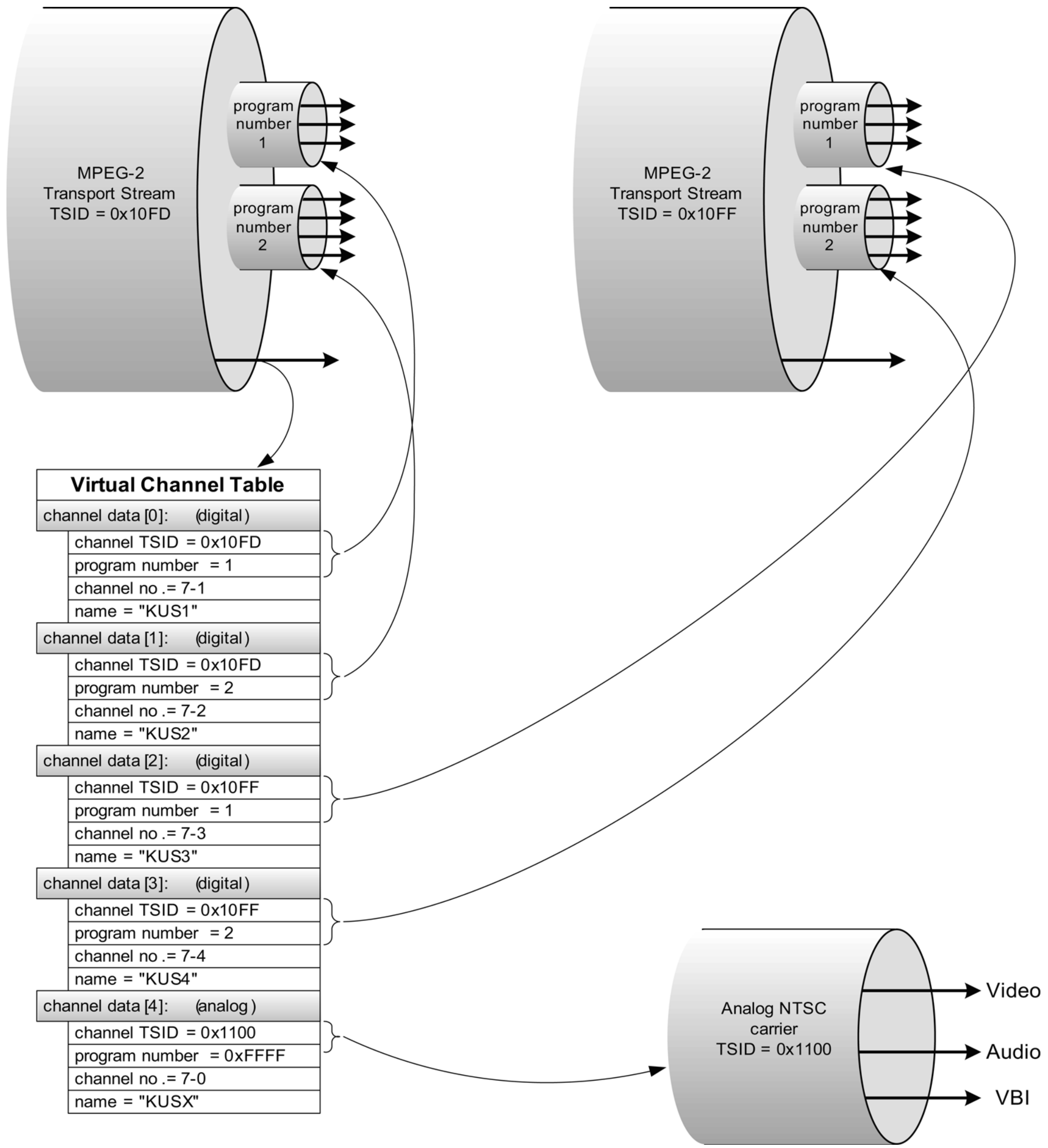

Fig. 8. VCT references. Adapted from: M. K. Eyer, *PSIP: Naming, Numbering, and Navigation for Digital Television* (New York: McGraw-Hill, 2003). Used with permission.

ETTs are linked to events announced in EITs by source ID (to identify the virtual channel) and by event ID (to reference a specific event on that channel). Yet another way ETTs are used is to deliver textual information regarding a virtual channel; this kind of ETT is called a channel ETT.

7) Overview of PSIP Tables: Fig. 9 illustrates a TS carrying PSIP tables. The ATSC PSIP standard defines PID value 0x1FFB as the "SI base PID" (another example of a fixed PID being used to speed channel acquisition). The packets identified by the base PID are shown at the top of the figure; they carry the STT, VCT, RRT, and MGT. Fig. 9 shows how the MGT references EIT and ETT table instances, indicating the PID values of the TS packets that carry each table section.

C. PSIP Descriptors

PSIP follows the MPEG-2 Systems standard in defining and making use of a number of "descriptors" to carry specific pieces of information. A descriptor is a simple data structure consisting of three basic parts: 1) an 8-bit "tag" value

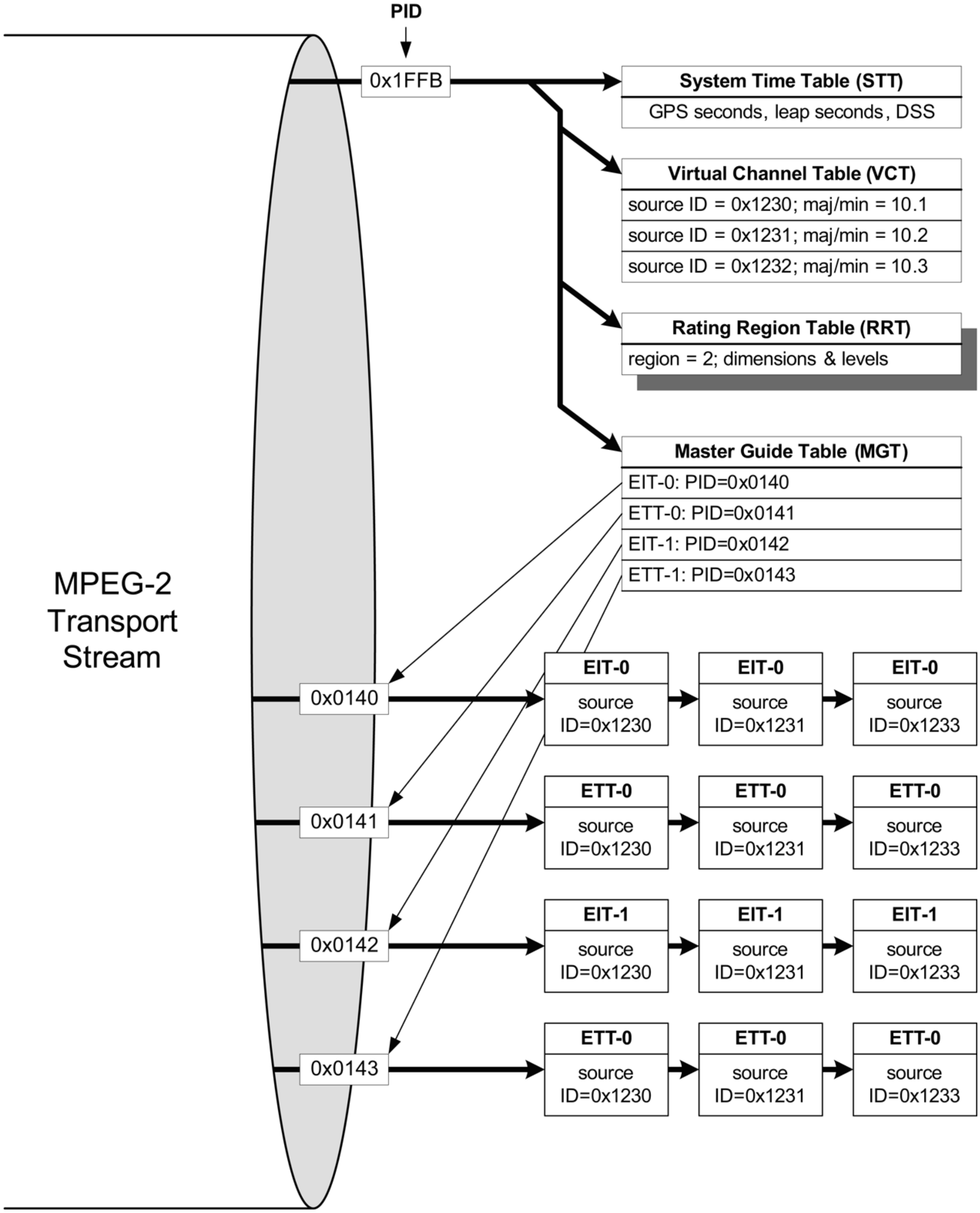

Fig. 9. The PSIP tables carried by a TS. Adapted from: Eyer, Mark K.: *PSIP: Naming, Numbering, and Navigation for Digital Television*, McGraw-Hill, New York, N.Y., 2003. Used with permission.

indicating the type of descriptor; 2) an 8-bit "length" field indicating the number of data bytes to follow; and 3) the data itself. Various MPEG-2 and PSIP table sections include syntactic locations where zero or more descriptors may be placed. Descriptors are context-sensitive. A descriptor of a given type may convey an entirely different meaning if found in location A as opposed to location B.

1) AC-3 Audio Stream Descriptor: One can learn the following information about an AC-3 audio stream from the AC-3 audio stream descriptor (see [5]):

- the sample rate of the encoded audio;
- an indication of the type of AC-3 coding employed in the stream;
- the bit rate (or upper bound of the bit rate) of the audio stream;
- the number of channels encoded in the audio stream;
- whether or not the audio stream is Dolby surround mode-encoded.

The AC-3 audio stream descriptor may also indicate the language of the associated audio track by a 3-byte ISO-639 language code field. The addition of the optional language bytes in the AC-3 descriptor allows the descriptor, when included in the EIT, to announce both the AC-3 stream characteristics and the language of each audio track that is offered with individual programming events. Audio language may also be indicated by an ISO-639 language descriptor, in accordance with MPEG-2 Systems guidelines.

2) ATSC Private Information Descriptor: One type of privately defined data in the TS comes in the form of a descriptor that might appear in a PMT section or in some other place where descriptors are allowed. The ATSC DTV standard specifies that the ATSC private information descriptor is to be used for the transport of private descriptor-based data. An entity wishing to include private information in a TS complying with the ATSC standard must first obtain a registered identifier assigned as a unique 32-bit number. Identifiers for the ATSC private information descriptor are issued by the SMPTE Registration Authority (SMPTE-RA). This `format_identifier` field is the same field used in the MPEG-2 registration descriptor (MRD).

3) Caption Service Descriptor: The function of the caption service descriptor is to announce the presence of line-21 compatibility bytes and/or DTV closed caption services carried with a given television program. Each block of caption service data is a fixed length and consists of an ISO 639 language identifier; a field indicating the type of caption service; a dual-purpose 6-bit field that identifies the caption service number for CEA-708-style captions (see [10]) or contains odd–even field information for the simulcast encapsulated ANSI/CEA-608-style captions (see [11]); a flag indicating whether or not the caption service is geared to younger readers; and finally, a flag that tells whether or not the caption service is formatted for video displays having a 16 : 9 aspect ratio.

4) Component Name Descriptor: Although the MPEG-2 Systems standard defines the PMT, which groups elementary stream components into MPEG-2 Programs, it does not define a way to associate a textual label with an elementary stream component of a Program. That is the function of the component name descriptor. Since the MPEG-2 Systems standard does not specify a standard method for encoding text strings for use in PSI or SI tables, the ATSC PSIP standard defines a multiple string structure for this purpose. The multiple string structure is an extension to the MPEG-2 Systems standard that defines a general data structure used specifically for text strings. It is used by the component name descriptor as well as other PSIP descriptors and tables.

The component name descriptor is helpful in cases where, for example, the Program includes more than one audio track of the same type and language. A sports broadcast could, for example have an English-language track for the "home team" announcers and a separate English-language track for the "visiting team" announcers. The textual name provided by the component name descriptor allows the receiver to create a user interface screen to ask the viewer which track to decode.

5) Content Advisory Descriptor: The content advisory descriptor is used to associate a given event with certain rating levels, e.g., extreme or mild, within certain dimentions, e.g., sexual content. The levels and dimensions are defined in a RRT that corresponds to a specific rating region, e.g., Canada. The descriptor is structured with a header followed by one or more data blocks containing the data specific to the particular rating region. The ATSC system for content advisories is described in Section III-E.

6) Extended Channel Name Descriptor: Channel names of at most seven characters are specified in the VCT in the `short_name` field. If a broadcaster like WXYZ wishes to multicast several SD channels, it can be difficult to retain the four-letter call sign and also differentiate the different channels when constrained to just seven characters. There are two ways to provide more information about a programming service. One way is to include an extended channel name descriptor. The other way is to include an ETT that is tied to the virtual channel. Typically, the ETT method is used to give a sentence or two of descriptive information about the channel, while the extended channel name descriptor method provides a longer channel name.

7) Redistribution Control Descriptor: The core purpose for the redistribution control descriptor (commonly referred to as the broadcast flag), is to signal that the content owner is asserting his or her rights with regard to redistribution of the program. The ATSC PSIP standard states that "The descriptor's existence within the ATSC stream shall mean: 'technological control of consumer redistribution is signaled.'" The FCC in the United States codified the broadcast flag in rules that may be found in 47 CFR 73.8 and 73.9 [7] and 47 CFR 76.1909 [12].

8) Service Location Descriptor: Decreasing the time between the moment a viewer enters a new channel number and the time decoded audio and video from the new channel can be presented has always been a goal of DTV systems engineering. Since the TVCT is always sent at a high repetition rate, it was considered desirable to replicate the PMT PID information in the TVCT to assist the receiver in more rapidly acquiring the selected Program. The service location descriptor provides a list of the service types and their corresponding PID values for the program elements that make up the programming service referenced by the virtual channel. It was further recognized that, since the PID assignments for the program elements within a given virtual channel are likely to remain relatively fixed and not change often, receivers can cache the data and use it to decrease channel acquisition time.

D. EPG

The EPG is probably the most widespread interactive television application in existence today. Viewers have become accustomed to pressing a single button on the remote control and seeing a grid, not unlike the newspaper listings, showing what television programs are currently on air, what television programs will be available in the future, extended descriptions of selected programs (events) and characteristics of those programs. Selecting a program from the grid will allow tuning to that program directly (if it is currently on the air) or in some circumstances automatically setting a digital video recorder (DVR) to record it. While the receiver designer determines the look and feel of the EPG, all of the information needed to populate the guide comes from PSIP

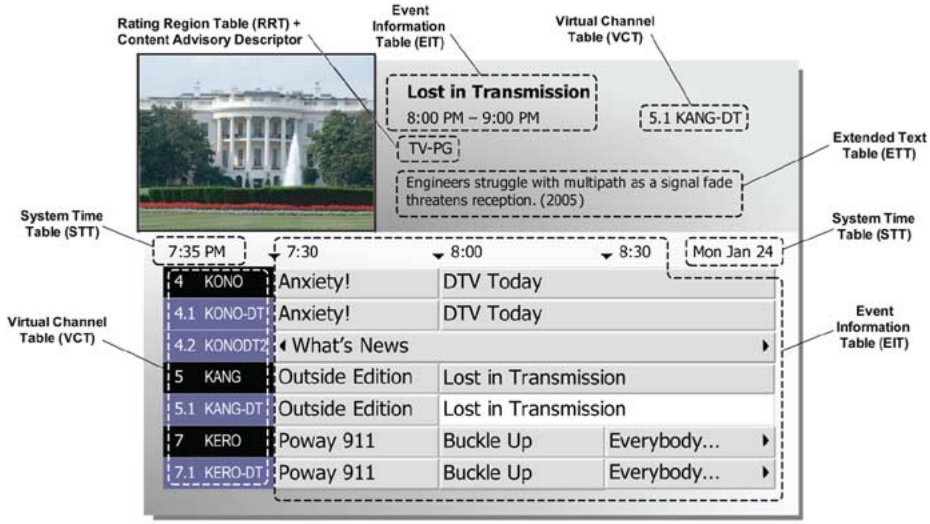

Fig. 10. EPG example.

tables in the ATSC TS. Receivers will extract the PSIP information from the available ATSC TSs and use it to build the database needed to implement the guide. Fig. 10 shows a typical EPG screen and which PSIP tables contribute to each element of the EPG display.

E. Content Advisory System

The ATSC PSIP standard defines the mechanism by which the V-chip content advisory data, now included in analog broadcasts in ANSI/CEA-608-B extended data service (XDS) packets (see [11]), is carried in an ATSC TS. A "content advisory" is information related to a given television program that tells something about the content of that program, such as whether or not it contains scenes of a violent or sexual nature. Although the V-chip system has very specific requirements for content labeling for use in the United States, PSIP was designed to accommodate the needs of any country in the world that adopts the ATSC DTV standard. PSIP defines a RRT that specifies the structure of a multidimensional content advisory system for a specific region, e.g., country, and a content advisory descriptor that can be used to associate specific program events with rating dimensions and levels defined in the RRT.

This section describes the structure and capabilities of the ATSC content advisory system.

1) Dimensions and Levels: The ATSC PSIP standard structures content advisory data using concepts called *dimensions* and *levels*. A "dimension" is a particular aspect of content or a particular way of characterizing content. Commonly used dimensions include the following.

- Age: This dimension specifies the minimum age recommended for viewers of the given content.
- Violence: The violence dimension indicates the prevalence or intensity of violent content.
- Sexuality: The sexuality dimension can indicate the presence of sexual situations, the explicitness of sexual scenes, or the adult nature of the program.
- Language: This dimension provides a way to characterize the extent of use of crude or profane language in the program.
- Standard rating system: It is possible to tie the definition of a rating dimension to an established rating system such as the scheme supported by the Motion Picture Association in the United States.

The term "level" is used to indicate the amount or intensity of the content, for a given dimension, that the program contains. For example, for the language dimension, three levels could be defined: 1) contains no offensive language; 2) contains some amount of bad language; or 3) contains profane language. Note that it is possible to include a rating level defined as "not rated for language." This "no information provided" rating is different from the rating level defined as "contains no offensive language."

Fig. 11 shows an example rating system consisting of four dimensions. As shown, rating dimensions are laid out in the horizontal direction while levels within each dimension are drawn vertically. Virtually any rating system may be mapped into a two-dimensional table such as this. Each cell in the table represents a specific rating level for a particular dimension. Referencing any cell in the table involves specifying the dimension (column) and level within that dimension (row).

2) Graduated Scales: Often, the definition of content ratings for a particular dimension are ordered in terms of increasing amount or intensity of the given content. The language example given above is an example of a dimension defined on a "graduated scale."

In the PSIP system, it is possible to define both rating dimensions that are defined on a graduated scale as well as those that are not. The advantage of using a graduated scale is that a receiver can be set up to block programming at a given level of content and to automatically know that blocking of content at higher levels is appropriate. For instance, to continue with the example language dimension, if the user were to decide to block programs with mildly of-

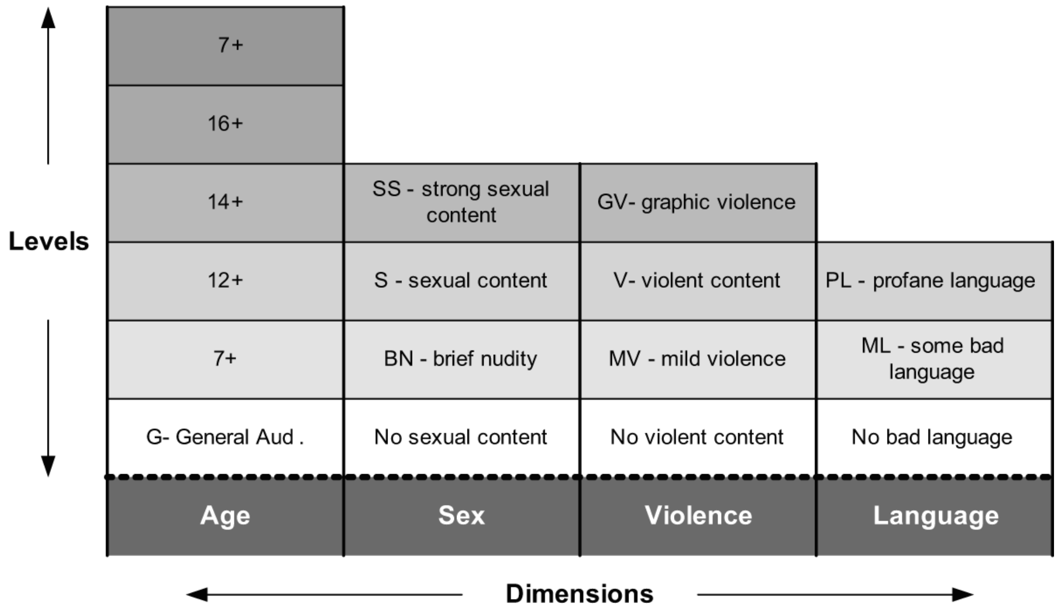

Fig. 11. Rating dimensions and levels in an example rating system.

fensive language, programming with profanity of any kind would also be blocked.

Consider the following example of an "age-based" rating dimension that is not defined on a graduated scale:

1) all age levels;
2) age 7 and up;
3) age 13 and up;
4) age 17 and up;
5) not rated.

Here, the "not rated" level may not really be an "age level" and thus may cause the content rating dimension not to work on a graduated scale. Assume that the system is defined such that a program rated "not rated" is not intended only for viewers that are, say, age 21 and up, but is given this rating simply because the program has not been reviewed and assigned an age rating, or perhaps because it was produced before the rating system came into existence.

In this case the viewer must be able to independently choose an age level at which to block and also whether or not to block programs rated "not rated." Since the "not rated" level in this example does not fit into a graduated scale, the receiver will need to be designed to treat it differently. In other words, if the viewer chooses to block programs rated "age 17 and up," or any lower age level, programs rated "not rated" will not be blocked unless the viewer specifically chooses to have them blocked as well.

3) System Design Goals: As mentioned, the PSIP content advisory system was designed with flexibility in mind. An important design goal was the concept that a receiving device could be built that would adapt and accommodate any content advisory system in use within any digital TS it happened to receive. It was felt that no prior knowledge about the system in use in a particular region or country should be needed. Another way to state this goal is to say that the definitions of the content rating system are "soft" rather than being hard-coded and inflexible.

Because the RRTs can be received and downloaded from within broadcast content, a receiver can be built to capture and to create from them an appropriate table-driven user interface. If a correction, revision, or even a complete change to an RRT were to be seen, a receiver's user interface could reflect that change automatically.

4) RRTs for the United States and Canada: The CEA-766-A standard [13] defines contents for RRTs corresponding to `rating_region` values 0x01 and 0x02. Value 0x01 defines the system for the United States and 0x02 the system for Canada. Two things become clear upon inspection of the RRT definitions for the United States. First and most importantly, the interpretation of the definitions of some of the dimensions for the United States region depend upon values in other dimensions. As one example, the "dialog" flag, when accompanying a TV-PG program means "some suggestive dialogue," but when that flag accompanies a TV-14 program, it means "intensely suggestive dialogue."

This interdependency means that a receiver cannot interpret the RRT for `rating_region` 0x01 simply by downloading the RRT and building a table-driven interface. Instead, the interdependencies described in the CEA standard must be factored into the design of the receiver.

CEA-766-A also defines an RRT for Canada. The RRT for Canada defines two dimensions, one for English-language ratings and the second for French-language ratings. Each dimension is a graduated scale, and there are no interdependencies. The CEA-766-A standard states that receiver implementations should be able to process updated versions of this RRT.

5) Rating a Program: Content advisory data pertinent to a particular program or event in PSIP is carried in a data structure called the content advisory descriptor. Each instance of a content advisory descriptor can provide rating information for up to eight regions. Within a given region it can indicate the rating level for any dimension defined in that region.

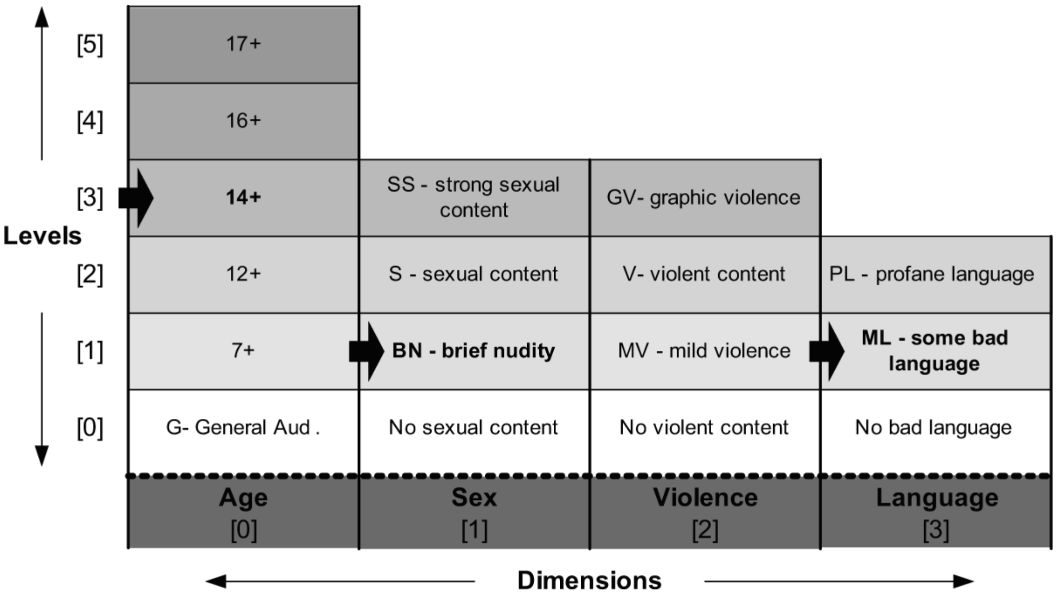

Fig. 12. Example program content advisory descriptor data.

Continuing with the example RRT definition given in Fig. 11, let us look at some example content advisory data. A content advisory (parental rating) for a television program is defined in terms of the RRT for a given region in which that program will be broadcast as illustrated in Fig. 12.

Fig. 12 shows an example of the data contained in content advisory descriptor for one region. Each content advisory descriptor provides data pertinent to a given program's rated content. In the example, the program is rated "14+" for age, BN for brief nudity, and ML because it contains some bad language. This program is not rated for violence content.

One program or event may be rated for more than one region. For example, a program may originate in the United States and be carried on Canadian or Mexican stations. In this case the content advisory descriptor would contain the separate data for the ratings according to each applicable RRT.

Data in the content advisory descriptor is formatted as a set of parameter pairs for each identified region. The parameter pairs each consist of a dimension index and a rating level. As shown in Fig. 12, one can consider the RRT to be a two-dimensional array, with dimensions in the X axis direction and levels in the Y axis direction.

Therefore, the content advisory descriptor data in the descriptor in the example of Fig. 12 would be encoded as: (0, 3), (1,1), and (3,1). This means dimension 0 is rated with a rating value of 3 (14+), dimension 1 is rated with a rating value of 1 (BN), dimension 2 is not rated, and dimension 3 is rated with a rating value of 1 (ML).

F. *Directed Channel Change (DCC)*

As discussed earlier, an ATSC TS can carry multiple virtual channels. There are a number of scenarios where it would be advantageous for the broadcaster to be able cause receivers to transparently tune (possibly without the viewer's knowledge) between virtual channels based on various criteria. These channel changes are "directed" by the broadcaster, thus the name DCC. The channel change may be to a virtual channel that is hidden or to a virtual channel that can be tuned normally. A number of selection criteria can be used in logical combinations to determine whether a channel change is to take place for a given receiver and if so, which channel to tune to. These criteria include location (zip code) and demographics. Examples might include: "viewers residing in zip code 98034 who are interested in football or basketball but not hockey or tennis" or "households in which no one works, some family members are under the age of 35, and no one is interested in comedy or musicals."

Two types of DCC are supported.

- Temporary retune, where the viewer will be switched to another virtual channel and then returned. An example might be switching viewers to the appropriate local news based on their location and then returning to the national news.
- Channel redirect, where the receiver will permanently tune to the new virtual channel. The channel redirect could be used when a station transitions from the broadcast of multiple SD channels to a single high-definition channel. Viewers watching one of the SD channels would be automatically tuned to the high-definition channel.

Fig. 13 gives an example timeline showing both types of transitions.

The elements of PSIP that support DCC are the DCCT and the DCC selection code table (DCCST). The DCCT indicates to the receiver that alternate programming is available, as well as the type, timing, and selection criteria for the change. The DCCST allows expansion of the selection codes defined in the ATSC PSIP standard, making the system extensible in the future.

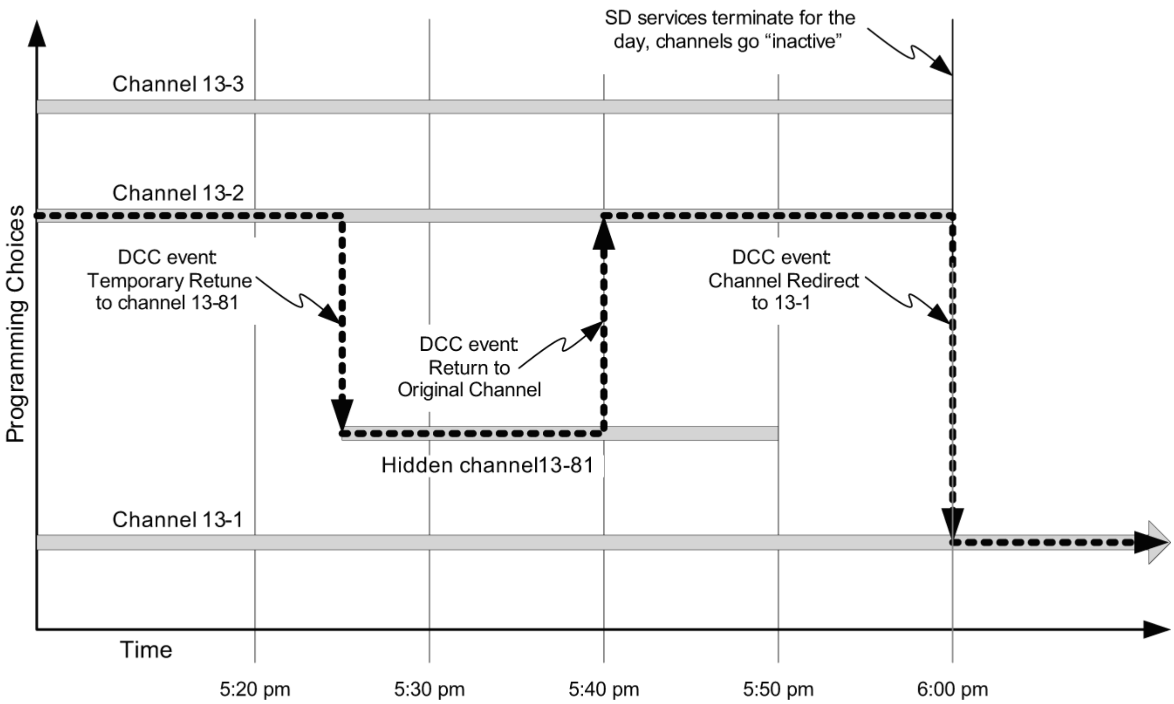

Fig. 13. Example DCC timeline.

IV. Private Data Transport

MPEG standards have numerous syntactical fields set aside for private use. When fields, tags, and table identifier fields are assigned a value by MPEG or by a standards body that is an MPEG user, such as ATSC, the values are then known as code points. In addition, many fields have ranges defined as user private. A user (not a standards body) is free to define one or more of these private fields, tags, and table values. Without some type of coordination mechanism, use of ATSC user private fields and ranges may lead to conflicts among privately defined services and metadata. Furthermore, without some form of scoping and registration, different organizations may inadvertently choose to use the same values for these fields, but with different meanings for the semantics of the information carried, causing confusion and inconsistent behavior for different receivers of the same broadcast TS. The ATSC DTV standard [1] places constraints on the use of private fields and ranges to avoid code point conflicts and to allow scoping of private data to the organization using it. The MRD is the main element that allows this scoping mechanism to be used.

The MRD (defined by MPEG-2 Systems) was designed to uniquely and unambiguously identify private data. The main content of the descriptor is a 4-byte format identifier. The Society of Motion Picture and Television Engineers (SMPTE) is the registration authority that assigns values for the format identifier field, guaranteeing that every assigned value will be unique. The ATSC DTV standard requires that any use of private data within the TS be scoped with the use of MRDs to avoid collisions.

Placement of the MRD in the PMT allows scoping to occur, as follows.

- When an MRD is located in the outer (Program level) loop of the PMT, then it signifies that any private information found within the entire MPEG-2 Program has meaning known only to the organization that registered the format identifier.
- When an MRD is located in an inner (program element) loop of the PMT, then it signifies that any private information found within the associated program element has meaning known only to the organization that registered the format identifier.

Only one MRD is allowed within any particular descriptor loop (if more than one were present, it would be unclear which has precedence). The presence of the MRD within a descriptor loop has meaning for the structure associated with the descriptor loop—not for any other descriptors within the same descriptor loop. The ATSC private information descriptor (see Section III-C2) uses the same format identifier field as the MRD and is used to carry any private descriptor information.

V. CA

ATSC Standard A/70A [14] defines the ATSC CA System for terrestrial broadcasting. The standard describes an architecture that is applicable to terrestrial broadcast systems and is not intended to apply to cable and satellite broadcast systems. The ATSC CA standard, based wherever possible on existing open standards, enables terrestrial broadcasters to implement CA services, e.g., pay-TV programs, using the ATSC DTV system. No business model is presumed (in fact, the standard allows multiple business models to coexist), but it is assumed that a CA system provider may exist as a separate entity from the broadcaster. The ATSC CA standard defines only those building blocks necessary to ensure interoperability between replaceable CA modules and CA-capable DTV receivers or set-top boxes, i.e., those designed to support the ATSC CA standard.

Compared to other ATSC standards, the ATSC CA standard is unique in that it does not describe precisely all of the techniques and methods needed to provide CA. Instead, it provides only the framework—the data envelopes and transport functions—that will allow several different CA systems to operate simultaneously and even compete with each other. In other words, different broadcasters may offer pay-TV services by means of different CA systems, each of which may have different transaction mechanisms, different security strategies, and more importantly, different business models. It is even possible for a single broadcaster to use two or more different CA systems for different pay-TV programs.

A. *System Elements*

The four basic elements of the ATSC CA system for terrestrial broadcast are: 1) the broadcast equipment; 2) the CA resources, i.e., scrambling keys and subscriber entitlement information, provided by the CA system provider; 3) the CA-capable DTV receiver or set-top box; and 4) the security module(s), also supplied by the CA system provider. The broadcast equipment generates the scrambled television programs and the necessary CA system information for over-the-air transmission to consumer receivers. The CA-capable DTV receiver demodulates the transmitted signals and passes the resulting TS to the security module for possible descrambling.

The ATSC CA standard specifies the use of triple-DES in cipher-block-chaining mode as the common scrambling algorithm for scrambling TS packets that contain content, e.g., video and/or audio elementary streams and/or data, of a pay-TV program. The standard also recommends that the interface between the broadcast equipment and the CA system provider comply with the Simulcrypt standard [15]. Simulcrypt is a DVB protocol defining equipment and methods for adequate information exchange and synchronization. The most important information elements that get exchanged are the scrambling keys.

The ATSC CA standard specifies the national renewable security standard (NRSS) [16] to define the interface between a DTV receiver and the replaceable security module. CA-capable DTV receivers must support either NRSS Part A (smart card) or NRSS Part B (PCMCIA) form factors or both. CA providers may distribute the plug-in NRSS security modules in a number of ways: directly, through consumer electronics manufacturers, through broadcasters, or through third-party agents. An NRSS module will typically contain information describing the program authorization status of the subscriber. Every time the NRSS module receives a TS from the DTV receiver with some of its Program components scrambled, the module will decide, based on its own information and system information in the TS, whether the subscriber is allowed access to one or more of those scrambled Programs. When the subscriber is allowed access, the packets of the selected Program are descrambled by the NRSS module one-by-one in real time and the resulting TS is passed back to the DTV receiver for decoding and display. For copy protection of the interface between the host DTV receiver and the security module, the ATSC CA standard relies on NRSS specifications.

B. *System Information for CA*

Besides carrying the scrambled signal, the TS carries PSI messages to identify properties and parameters of the scrambled signals. The required information is, as defined in the MPEG-2 Systems standard [2]: the CAT; entitlement control messages (ECMs); EMMs; and two CA descriptors, one of which is defined by the ATSC CA standard. The CA system provider in general, privately defines the syntax of the ECM and EMM content. ECMs are data units that carry the key for descrambling the signals. EMMs provide general information to subscribers and usually contain information about the status of the subscription itself. Broadcasters interested in providing conditionally accessed services through one or more CA providers need to transmit ECMs and EMMs for each of those CA providers. An NRSS module must be able to understand the privately defined EMMs and ECMs for the Programs it is authorized to descramble.

The terrestrial broadcast TS carries an EPG as defined in the ATSC PSIP standard and described in Section III. The program guide contains detailed information about present and future events that may be useful for the implementation of a CA system. For this reason, The ATSC CA standard defines a descriptor called the `ATSC_CA_descriptor()` that can be placed in either the VCT or the EIT of PSIP. Similar to the definitions of ECMs and EMMs, only the generic descriptor structure is defined while its content is private. The CA-capable DTV receiver is required to parse the program guide tables in search of this descriptor and pass it to the NRSS module. Due to its private content, it is the NRSS module that ultimately processes the information. Note that although PSIP does not require use of CA, the ATSC CA standard does require the use of PSIP.

VI. TRANSPORT SUBSYSTEM BITSTREAM INTERFACES

Generation of a TS compliant with the ATSC DTV standard [1] requires a physical hardware/software transport subsystem. The transport subsystem receives as inputs the various MPEG-2 Programs and encapsulated data streams including PSIP data and provides the compliant TS as its output. The ATSC DTV standard does not specify specific physical interfaces between facility equipment. However, to ensure interoperability among equipment from different suppliers various voluntary industry interface standards are generally followed.

A. *Transport Subsystem Input Characteristics*

The MPEG-2 Systems standard [2] defines the input to the transport subsystem as MPEG-2 elementary streams. However, specific physical implementations may include the PES packetizer within a video, audio, or other data encoder. Also common industry practice and the ATSC DTV standard consider `private_section` encapsulated data to be a parallel layer to PES and again a `private_section` encapsulator may be included within a data encoder; and not as part of the transport subsystem. Therefore, the inputs to the transport subsystem may be elementary streams, PES packets, or `private_section` encapsulated data.

B. Transport Subsystem Output Characteristics

Conceptually, the output from the transport subsystem is a continuous MPEG-2 TS at a constant rate of T_r Mb/s when transmitted in an ATSC 8 VSB system and $2T_r$ when transmitted in a 16 VSB system where

$$T_r = 2 \times \left(\frac{188}{208}\right)\left(\frac{312}{313}\right)\left(\frac{684}{286}\right) \times 4.5 = 19.39\ldots \text{ Mb/s}$$

The symbol rate S_r in Megasymbols per second for the ATSC transmission subsystem is given by:

$$S_r = \left(\frac{684}{286}\right) \times 4.5 = 10.76\ldots \text{Msymbols/s.}$$

See Annex D of [1] for a complete explanation of the factors in the above equations. The ATSC DTV standard requires that T_r and S_r be locked to each other in frequency.

VII. Summary and Conclusions

The transport layer of the ATSC DTV system has been described. The constraints placed on the MPEG-2 Systems standard by the ATSC DTV standard have been explained. The extensions to the MPEG-2 Systems standard defined by the ATSC DTV standard, the ATSC PSIP standard, and the ATSC CA standard have also been covered. It has been shown how the ATSC standards use the MPEG-2 TS toolkit to meet the demanding requirements of a terrestrial broadcast DTV system, particularly by adding the comprehensive PSIP extension, which enables user-friendly navigation and provides data from which a receiver can construct an EPG. The inherent flexibility of the ATSC transport layer to accommodate new requirements is already being exercised. Backward-compatible extensions and enhancements to add additional features to the ATSC system, such as providing transport layer support for Enhanced VSB (E-VSB) video and audio coding, are currently under development within ATSC. Additional extensions can be expected in the future.

References

[1] *ATSC Digital Television Standard*, ATSC A/53D, Advanced Television Systems Committee, Washington, DC, Jul. 19, 2005 [Online]. Available: http://www.atsc.org

[2] Information technology—Generic coding of moving pictures and associated audio—Part 1: Systems , ISO/IEC 13818-1:2005, International Telecommunications Union [Online]. Available: http://www.itu.org, as ITU-T Rec. H.222.0, Global Engineering Documents [Online]. Available: http://www.global.ihs.com

[3] *Program and System Information Protocol for Terrestrial Broadcast and Cable (PSIP)*, ATSC A/65B, Advanced Television Systems Committee, Washington, DC, Mar. 18, 2003 [Online]. Available: http://www.atsc.org

[4] Information technology—Generic coding of moving pictures and associated audio—Part 2: Video , ISO/IEC 13818-2:2002, International Telecommunications Union [Online]. Available: http://www.itu.org, (as ITU-T Rec. H.222.0), Global Engineering Documents [Online]. Available: http://www.global.ihs.com

[5] *Digital Audio Compression (AC-3, E-AC-3) Standard*, ATSC A/52B, Advanced Television Systems Committee, Washington, DC, Jun. 14, 2005 [Online]. Available: http://www.atsc.org

[6] R. J. Crinon, D. Bhat, D. Catapano, G. Thomas, J. Van Loo, and G. Bang, "Data broadcasting and interactive television," *Proc. IEEE,* vol. 94, no. 1, pp. 102–118, Jan. 2006.

[7] Rules of the U.S. Federal Communications Commission, 47 CFR Part 73, Washington, DC [Online]. Available: http://wireless.fcc.gov/rules.html

[8] *Cable Television Channel Identification Plan*, CEA-542-B, Consumer Electronics Association, Arlington, VA, Jul. 2003 [Online]. Available: http://www.ce.org [Online]. Available: http://www.ce.org, Global Engineering Documents [Online]. Available: http://www.global.ihs.com

[9] *Data Broadcast Standard with Amendment 1 and Corrigendum 1 and 2*, ATSC A/90, Advanced Television Systems Committee, Washington, DC, Jul. 26, 2000 [Online]. Available: http://www.atsc.org

[10] *Digital Television (DTV) Closed Captioning*, ANSI/CEA-708-B, Consumer Electronics Association, Arlington, VA, Jun. 1997 [Online]. Available: http://www.ce.org, Global Engineering Documents [Online]. Available: http://www.global.ihs.com

[11] *Line 21 Data Services*, ANSI/CEA-608-C, American National Standards Institute and Consumer Electronics Association, Aug. 2005 [Online]. Available: http://www.ce.org, Global Engineering Documents [Online]. Available: http://www.global.ihs.com

[12] Rules of the U.S. Federal Communications Commission, 47 CFR Part 76, Washington, DC [Online]. Available: http://wireless.fcc.gov/rules.html

[13] *U.S. and Canadian Rating Region Tables (RRT) and Content Advisory Descriptors for Transport of Content Advisory Information Using ATSC A/65A Program and System Information Protocol (PSIP)*, CEA-766-A, Consumer Electronics Association, Apr. 2001 [Online]. Available: http://www.ce.org/, Global Engineering Documents [Online]. Available: http://www.global.ihs.com

[14] *Conditional Access System for Terrestrial Broadcast*, ATSC A/70A, Advanced Television Systems Committee, Washington, D.C., Jul. 22, 2004 [Online]. Available: http://www.atsc.org

[15] *Technical Specification of DVB Simulcrypt*, ETSI TS 101 197 V1.2.1 (2002-02), European Telecommunications Standards Institute, Feb. 2002 [Online]. Available: http://www.etsi.org

[16] *National Renewable Security Standard (NRSS)*, CEA 679-C, Consumer Electronics Association, Mar. 2000 [Online]. Available: http://www.ce.org, Global Engineering Documents [Online]. Available: http://www.global.ihs.com

Bernard J. Lechner (Life Fellow, IEEE) received the B.S.E.E. degree from Columbia University, New York, and has done graduate work at Princeton University, Princeton, NJ, and the Harvard School of Business, Cambridge, MA.

He is one of the world's leading experts on television and display systems. He was extensively involved in technical research on advanced television and display systems and in the development of standards for High Definition Television (HDTV). He has consulted to government and industry on all aspects of television and display systems. His clients include Fortune-5O0 companies, as well as a number of start-up companies in the television and display industries. Formerly Staff Vice President, Advanced Video Systems, RCA Laboratories, his 30-year career at RCA covered all aspects of television and display research, from early work on home video tape recorders in the late 1950s, extensive development of flat-panel matrix displays in the 1960s including pioneering efforts on active-matrix liquid crystal displays, advanced two-way cable TV systems and pay-TV systems in the early 1970s, electronic tuning systems and CCD comb-filters for TV

receivers in the mid-1970s, automated broadcast cameras and CCD broadcast cameras in the late 1970s and early 1980s, to HDTV in the mid-1980s. The National Academy of Television Arts and Sciences awarded Emmys to two of the broadcast camera projects for which he led the research team. He received two RCA Laboratories Outstanding Achievement Awards and a David Sarnoff Team Award in Science. He was an active participant in the various groups working on standards for advanced television systems in the United States, including the FCC Advisory Committee on Advanced Television Service (ACATS), the Society of Motion Picture and Television Engineers (SMPTE), and the Advanced Television Systems Committee (ATSC). During 1989 and 1990, he served as a member of the United States delegation to the extraordinary and final meetings of the Comite Consultatif International des Radiocommunications (CCIR) in Geneva concerning international HDTV standards. He was chairman of the ATSC Specialists Group on digital TV transport standards from its inception in 1994 until April 2002, and he also chaired the CEA/NCTA Digital Standards Subcommittee relating to standards for "cable-ready" digital television receivers. He was chairman of the Teletext Committee of the Electronic Industries Association from 1980 to 1986 and was a member of the National Cable Telecommunications Association Engineering Committee from 1977 to 2002. He served for 15 years as Chairman of the Advisory Commission for Electrical Engineering at Mercer County Community College and was a member of the Board of Directors of Palisades Institute for Research Services from 1981 to 2002. He holds ten U.S. patents and is widely published in the areas of displays and television systems.

Mr. Lechner is a Life Fellow of the Society for Information Display (SID) and the SMPTE. He is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi. In 1971, SID named him the first recipient of the Frances Rice Darne Award for his outstanding contributions to matrix displays. In 1972, he was elected to the SID Board of Directors and subsequently served as Treasurer, Secretary, Vice President, and President (1978–1980) of SID. In 1983 he was named the first recipient of the Beatrice Winner Award for his contributions to SID. In 1996, he was awarded the David Sarnoff Gold Medal by SMPTE for his many contributions to the technologies essential to today's television systems. In May 2000, he was honored by the ATSC as the first recipient of the ATSC Outstanding Contributor Award. The award has become an ATSC annual award named in his honor. In November 2001, he received the SMPTE Progress Medal Award, SMPTE's highest honor, which was presented to Mr. Lechner in recognition of his outstanding technical contributions to the progress of the engineering phases of the television industry. In April 2002, the National Association of Broadcasters (NAB) presented him the NAB Television Engineering Achievement Award. The Award, NAB's highest technical honor, was given to him in recognition of his outstanding contributions to the development of advanced television and display systems and for his leadership in the development of cross-industry digital television standards.

Richard Chernock (Member, IEEE) received the Sc.D. degree in nuclear materials engineering from the Massachusetts Institute of Technology (MIT), Cambridge

He was a Research Staff Member at IBM Research, investigating digital broadcast technologies. At IBM, he used transmission electron microscopy to study materials characteristics for advanced ceramics packaging and semiconductor technology. He is currently Director of Technology, Triveni Digital, Princeton Junction, NJ. In that position, he is developing strategic directions for metadata management, content distribution, and monitoring in emerging digital television systems and infrastructures.

Dr. Chernock is active in a number of Advanced Television Systems Committee (ATSC) standards committees, particularly in the area of metadata and data broadcast. He chairs a number of *ad hoc* committees within ATSC whose work relates to metadata and transport issues. He is vice-chair of the Technology Group on Distribution (TSG).

Mark K. Eyer received the B.S. degree (*cum laude*) and the M.S.E.E. degree from the University of Washington,Seattle, in 1973 and 1978, respectively.

Since 1994, he has been involved in the standards development process and made contributions to various digital television standards published by the Consumer Electronics Association, Society for Cable Telecommunications Engineers, and the Advanced Television Systems Committee (ATSC). He is currently Director of Systems for the Technology Standards Office of Sony Electronics, Bellevue, WA. He represents Sony in various standards committees in the United States and contributes systems engineering expertise to development of Sony's digital television products. His book *PSIP: Program and System Information Protocol* (McGraw-Hill, 2002) is based on his work on the ATSC A/65 standard. For the past 20 years, he has been involved with the development of technologies and products related to secure and digital television and he holds 18 U.S. patents in these areas.

Mr. Eyer received the 2005 Finegan Standards Medal from the American National Standards Institute (ANSI). This award honors an individual who has shown extraordinary leadership in the actual development and application of voluntary standards.

Adam Goldberg (Senior Member, IEEE) received the B.S. degree in computer science from Iowa State University, Ames, in 1992.

Previously, he held positions at C-Cube Microsystems, DiviCom, and Microware Systems. He is currently the Director, Television Standards and Policy Development at Sharp Laboratories of America, Fairfax, VA. In that role, he is Sharp's primary representative to television-related standards-making activities in the United States. He is also the current Chair of the CEA R4.3 committee and was the Chair of the CEA working group that developed a recommended practice for PSIP implementation for receivers. He has extensive experience in digital television and particularly MPEG-2 Systems (ISO/IEC 13818-1) and its applications (like PSIP), and has been involved in ATSC, DVB, and SCTE engineering committees.

Mr. Goldberg is a member of the Society of Motion Picture and Television Engineers (SMPTE) and the Society of Cable Telecommunications Engineers (SCTE).

Matthew S. Goldman (Senior Member, IEEE) received the B.S. (high honors) and M.S. degrees in electrical engineering from Worcester Polytechnic Institute, Worcester, MA, in 1983 and 1988, respectively.

He has been actively involved in the development of digital television systems since 1992. Until 1996, he was a consulting engineer at Digital Equipment Corporation, where he was the systems architect for a first generation digital video server system. From 1996 to 2000, he was director of engineering, advanced systems development, at DiviCom Inc, where he specified MPEG-2 based systems solutions. He is currently Vice President of Technology, Compression Systems, for TANDBERG Television Inc., Bedford, NH. He was a prominent participant and a project editor of the MPEG-2 standards, and has been influential in other industry organizations, including the Society of Cable Telecommunications Engineers, the Advanced Television Systems Committee, and the Society of Motion Picture and Television Engineers (SMPTE). He holds six patents related to digital video transport.

Mr. Goldman is a member of the SMPTE and a member of the Academy of Digital Television Pioneers.