IP SHOWCASE™ THEATER

CURATED BY VSF VIDEO SERVICES FORUM

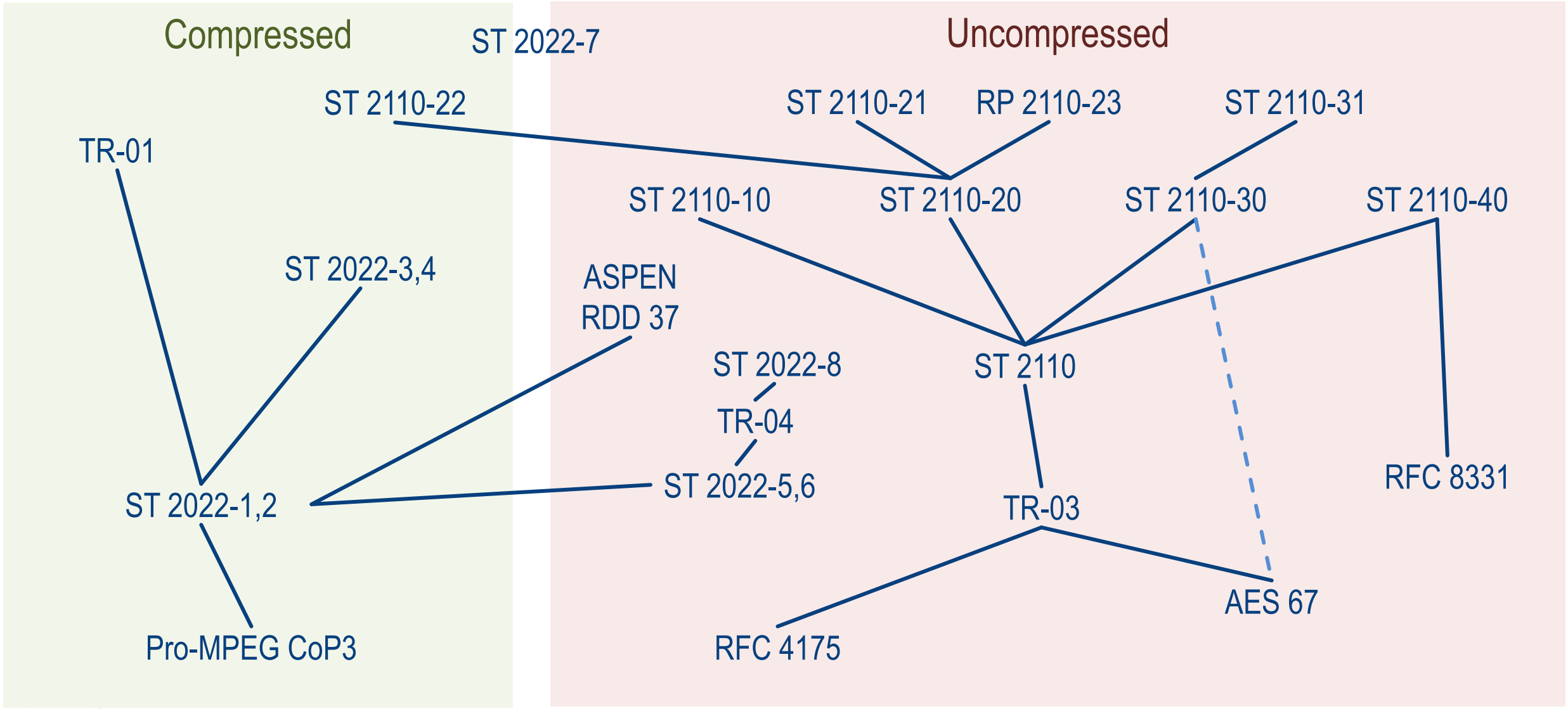
SMPTE ST 2110 in 60 Minutes

Wes Simpson

President, Telecom Product Consulting

telecompro.tv
IP Video Training and Consulting

IP Video Evolutionary Tree



IP SHOWCASE™ THEATER

Elements of ST 2110

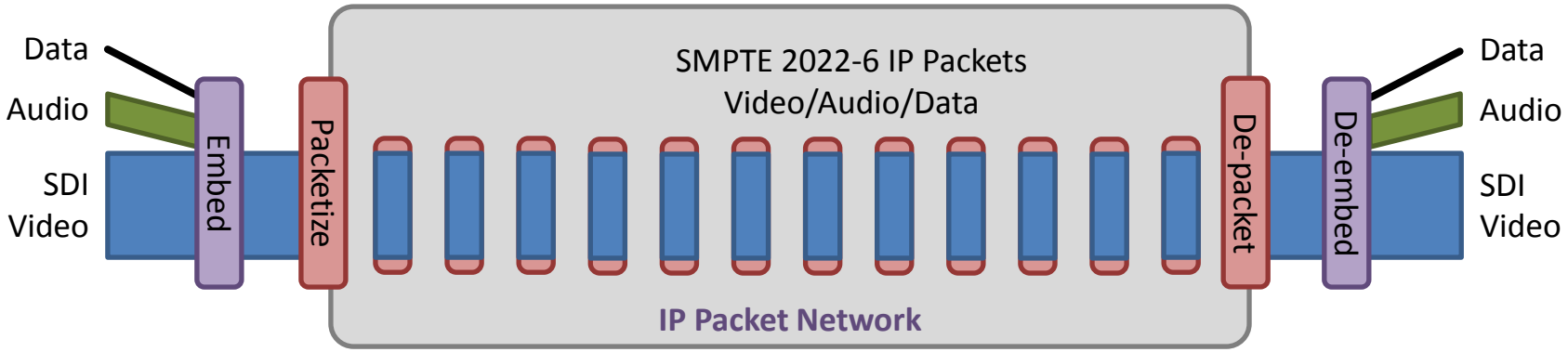
- ST 2110-10 System and Timing
- ST 2110-20 Uncompressed Video
- ST 2110-21 Video Stream Packet Shaping
- ST 2110-30 Uncompressed Audio
- ST 2110-31 AES3 Audio Streams
- ST 2110-40 Ancillary Data

New Elements of ST 2110

- OV 2110-0 Roadmap for the 2110 Document Suite
- ST 2110-10 System and Timing
- ST 2110-20 Uncompressed Video
- ST 2110-21 Video Stream Packet Shaping
- ST 2110-22 Constant Bit-Rate Compressed Video
- RP 2110-23 Single Video Essence Transport over Multiple ST 2110-20 Streams
- ST 2110-30 Uncompressed Audio
- ST 2110-31 AES3 Audio Streams
- ST 2110-40 Ancillary Data

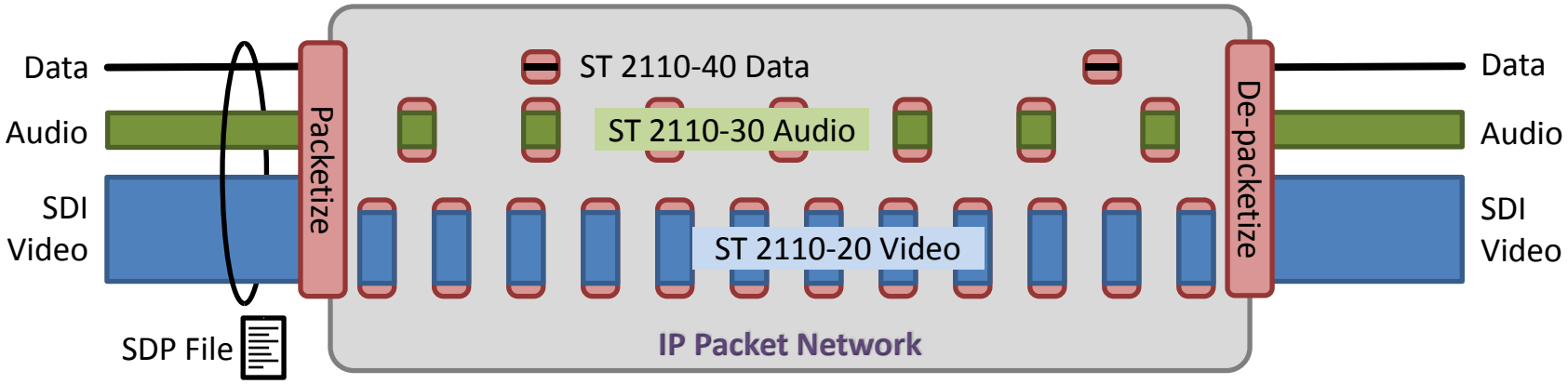
ST 2022-6 High Bit Rate Media over IP

- Take entire SDI signal and encapsulate it in IP stream
  - Includes audio and embedded data signals
- Easy to maintain audio/video synchronization
  - Hard to process just one part of a stream



ST 2110

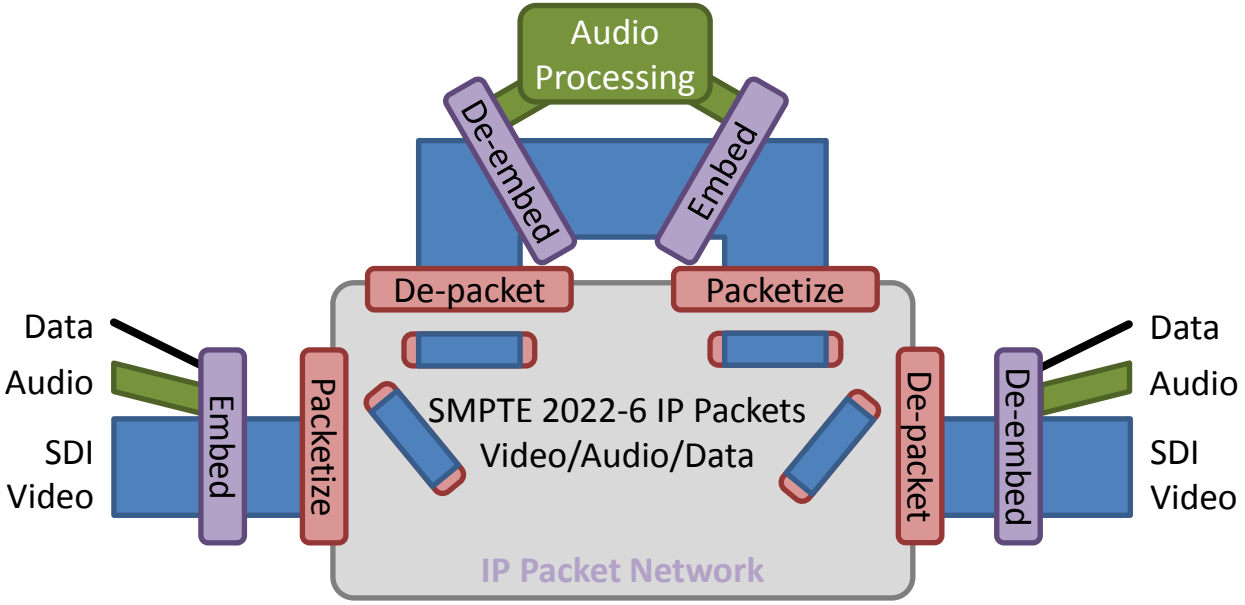
- Each media type in a separate packet stream
  - Easy to process individual components
  - Signals need to be resynchronized after processing
- PTP (Precision Time Protocol) used for packet timestamping



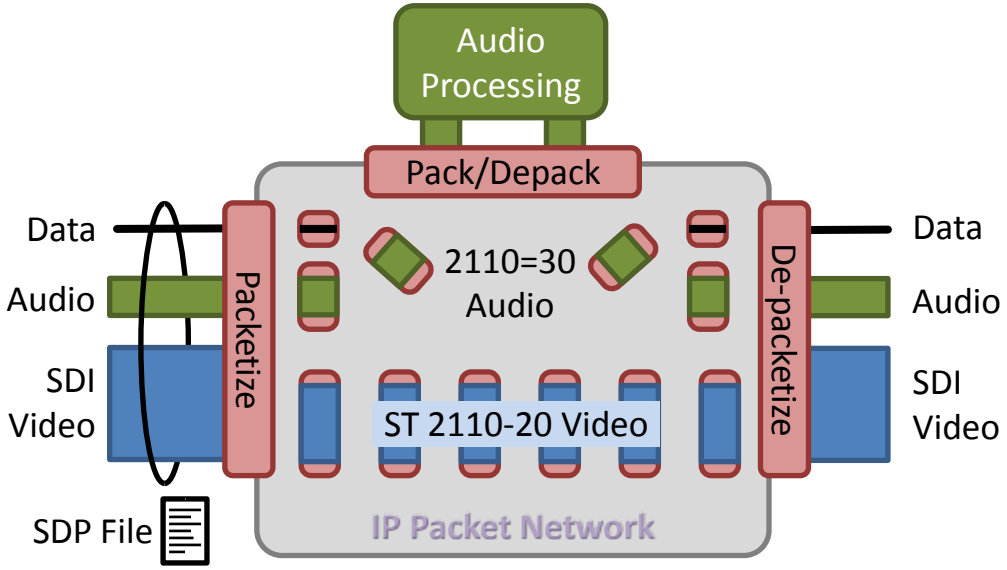
IP SHOWCASE™ THEATER

Audio Processing Packet Flow

Using SDI/ST 2022-6



Using ST 2110

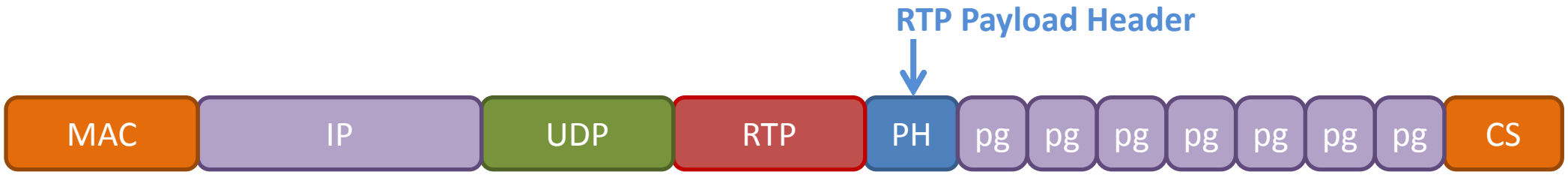


telecompro.tv
IP Video Training and Consulting

ST 2110-10 System Timing and Definitions

- Maximum UDP datagram size: 1460 octets, including UDP header
  - Extended UDP datagram allowed with up to 8960 octets
- SMPTE ST 2059-2 PTP Profile of IEEE 1588-2008
  - If interchanging audio with AES67, then compatible parameters must be used
- RTP timestamps are tied to the media
  - For video, RTP timestamps of all packets for video frame are the same
  - For real-time sources, this should represent the Image Capture Time
  - For SDI converters, RTP timestamp is moment when video frame alignment point arrives at device input (SMPTE ST2059-1 defines alignment points)
- All media clocks must have an offset of zero
  - This makes it easier to recover from loss of signal or unexpected system restart

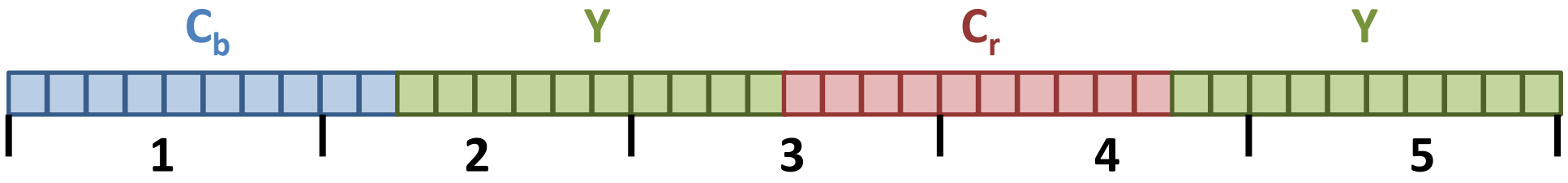
ST 2110-20 Video Encapsulation



- Multiple video pixel groups (pgroups)
- RTP Payload Header applied
- Inserted into an RTP packet
- Placed into UDP packet
- IP packet header attached
- Wrapped into Ethernet Frame

ST 2110-20 Pixel Groups

- Pixels formed into pgroups
  - pgroup size depends on sampling format
  - Must be integer number of octets
  - Pixels that share samples must be in the same pgroup
- Example: 4:2:2 10-bit
  - 2 pixels in 5 octets

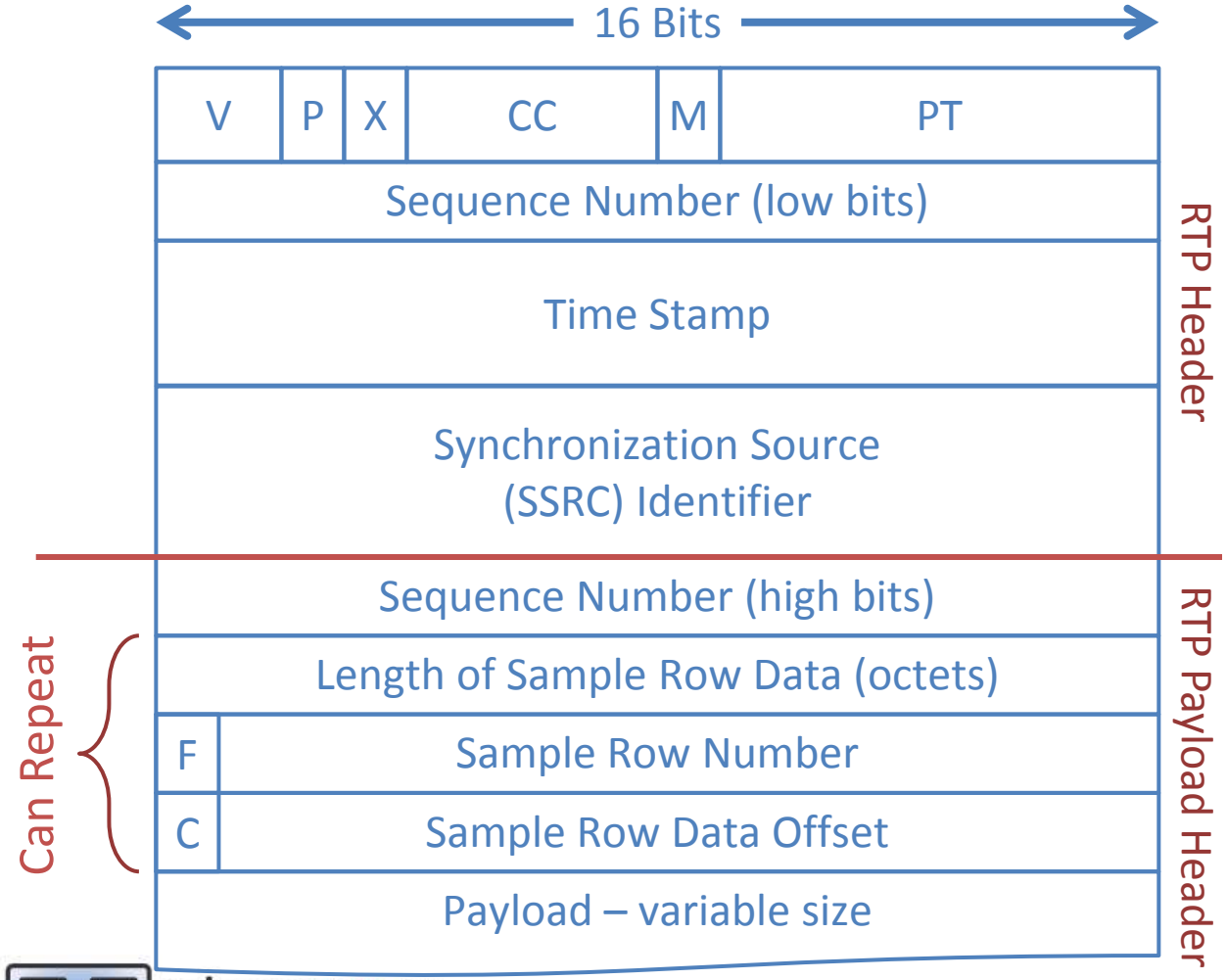


Pixel Group Sizes

- Every supported video format listed in ST 2110-20 tables
  - Tables also include order of samples within each pgroup

sampling	depth	pgroup size (octets)	pgroup coverage (pixels)	Sample Order
YCbCr-4:2:2 CLYCbCr-4:2:2	8	4	2	C'_B,Y0',C'_R,Y1'
	10	5	2	C'_B,Y0',C'_R,Y1'
	12	6	2	C'_B,Y0',C'_R,Y1'
	16, 16f	8	2	C'_B,Y0',C'_R,Y1'
ICtCp-4:2:2	8	4	2	C'_T,I0',C'_P,I1'
	10	5	2	C'_T,I0',C'_P,I1'
	12	6	2	C'_T,I0',C'_P,I1'
	16, 16f	8	2	C'_T,I0',C'_P,I1'

ST 2110-20 Video Packet Header



- 32-bit Sequence Number (16 bit Sequence number would wrap in less than half a second for Gigabit-class payloads)
- Length of Sample Row Data = Number of octets from scan line in this datagram. Must be multiple of pgroup
- F = 0 for progressive scan and first field in interlace video
- F = 1 for second field in interlace video
- Video Line Number = Video scan line number, starts at 0 for first active line of video (note difference from SDI line numbering)
- C = 1 if more than one line is in datagram, set to 0 for last line in each datagram
- Sample Row Data Offset
  - = Location of first pixel of payload data within scan line
  - = 0 if first pixel in scan line; counts by pixels

Calculating Video Stream Packet and Bit Rates

- Step 1: Gather data about video signal:
  - Image Size (image height in lines, image width in pixels)
  - Sampling system (e.g YCbCr-4:2:2) and sample depth (e.g. 10 bits)
  - Frame Rate

Image Size: 1920x1080 Sampling: 4:2:2 10-bit Frame Rate: 59.94

More Calculating

- Step 2: Figure out RTP payload size in bytes and pixels
  - Per ST 2110, standard MAXUDP is 1460 bytes
  - UDP and RTP headers are 8 bytes and 12 bytes, for a total of 20 bytes
  - Worst case datagram with pixels from two rows: RTP Payload Header of 14 bytes
  - Subtract headers from MAXUDP to get available RTP payload

Image Size: 1920x1080 Sampling: 4:2:2 10-bit Frame Rate: 59.94

UDP Payload = 1460 - 8 (UDP) - 12(RTP) - 14 (Payload Header) = 1426 bytes Payload

And Some More Calculating

- Step 3: Calculate max number of pgroups and pixels in a packet
  - Using ST 2110-20 table, select correct pgroup size in bytes and pixels
  - Divide available payload by pgroup size
  - Take result and round down (truncate) – can't have a partial pgroup
  - Multiply pgroups/packet by size of pgroup in pixels to get pixels per packet

Image Size: 1920x1080 Sampling: 4:2:2 10-bit Frame Rate: 59.94

UDP Payload = 1460 - 8 (UDP) - 12(RTP) - 14 (Payload Header) = 1426 bytes Payload

1426 bytes / 5 (bytes/pgroup) = 285.2, round DOWN to 285 pgroups/packet

285 pgroups x 2 (pixels/pgroup) = 570 (pixels/packet)

The Calculations Continue…

- Step 4: Determine number of packets per video frame
  - Multiply image width by height to get total pixels in each frame
  - Divide total pixels by pixels per packet to get packets per frame
  - Note: Must round result of this division up – last packet may be partially filled

Image Size: 1920x1080 Sampling: 4:2:2 10-bit Frame Rate: 59.94
UDP Payload = 1460 - 8 (UDP) - 12(RTP) - 14 (Payload Header) = 1426 bytes Payload
1426 bytes / 5 (bytes/pgroup) = 285.2, round DOWN to 285 pgroups/packet
285 pgroups x 2 (pixels/pgroup) = 570 (pixels/packet)
1920 pixels x 1080 lines = 2,073,600 pixels/frame
2,073,600 / 570 (pixels/packet) = 3637.89 round UP to 3638 packets/frame

Even More Calculating

- Step 5: Calculate total size of each packet on wire
  - Multiply number of pgroups per packet by size of pgroup in bytes
  - Determine UDP packet size: (payload in bytes) + 14 + 12 + 8
  - Add IP header (20), Ethernet + VLAN (22), plus preamble and gap (20)
  - Total is size of each packet on wire

Image Size: 1920x1080 Sampling: 4:2:2 10-bit Frame Rate: 59.94

UDP Payload = 1460 - 8 (UDP) - 12(RTP) - 14 (Payload Header) = 1426 bytes Payload

1426 bytes / 5 (bytes/pgroup) = 285.2, round DOWN to 285 pgroups/packet

285 pgroups x 2 (pixels/pgroup) = 570 (pixels/packet)

1920 pixels x 1080 lines = 2,073,600 pixels/frame

2,073,600 / 570 (pixels/packet) = 3637.89 round UP to 3638 packets/frame

285 pgroups x 5 (bytes/pgroup) + 14 + 12 + 8+ 20 (IP) + 22 (VLAN) + 20 = 1521 bytes/pkt.

The Final Step

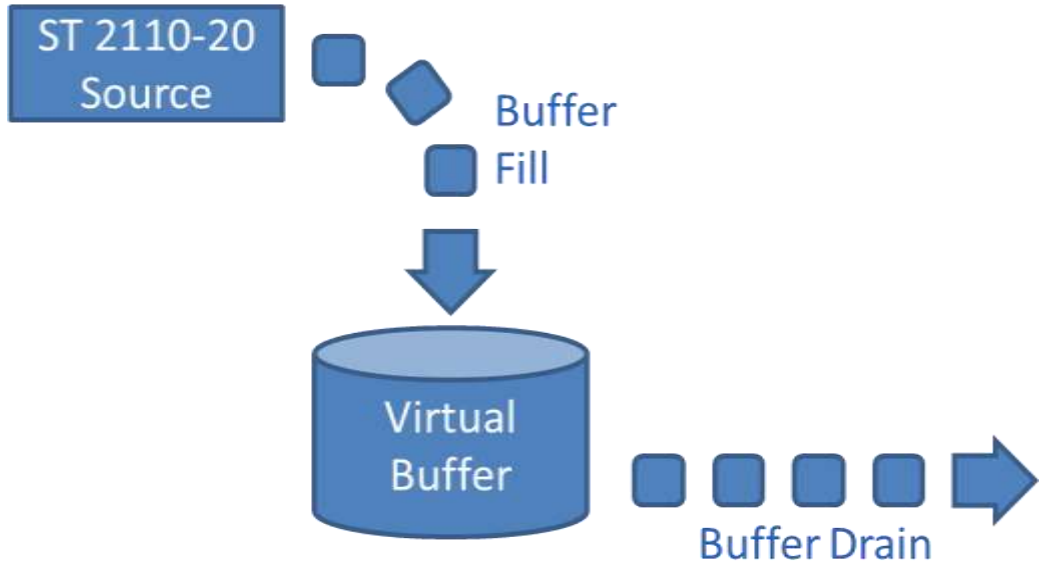
- Step 6: Determine stream bit rate
  - Multiply packets/frame by frame rate of signal (ok to not round)
  - Multiply packets per second by bytes per packet
  - Multiply by 8 to convert bytes to bits per second
  - Express final result in Gigabits per second

Image Size: 1920x1080 Sampling: 4:2:2 10-bit Frame Rate: 59.94
UDP Payload = 1460 - 8 (UDP) - 12(RTP)- 14 (Pay. Head.) = 1426 bytes
1426 bytes / 5 (bytes/pgroup) = 285.2, round DOWN to 285 pgroups/packet
285 pgroups x 2 (pixels/pgroup) = 570 (pixels/packet)
1920 pixels x 1080 lines = 2,073,600 pixels/frame
2,073,600 / 570 (pixels/packet) = 3637.89 round UP to 3638 packets/frame
285 pgroups x 5 (bytes/pgroup) + 14 + 12 + 8 + 20 (IP) + 22 (VLAN) + 20 = 1521 bytes/pkt.

3638 packets/frame x 59.94 x 1521 x 8 = 2.65 Gbit/s

ST 2110-21 Timing Models

- Senders can’t burst out all of their data at once
  - Overloads receivers and network switch buffers
- Some variability is necessary
  - HANC/VANC gaps, software-based senders

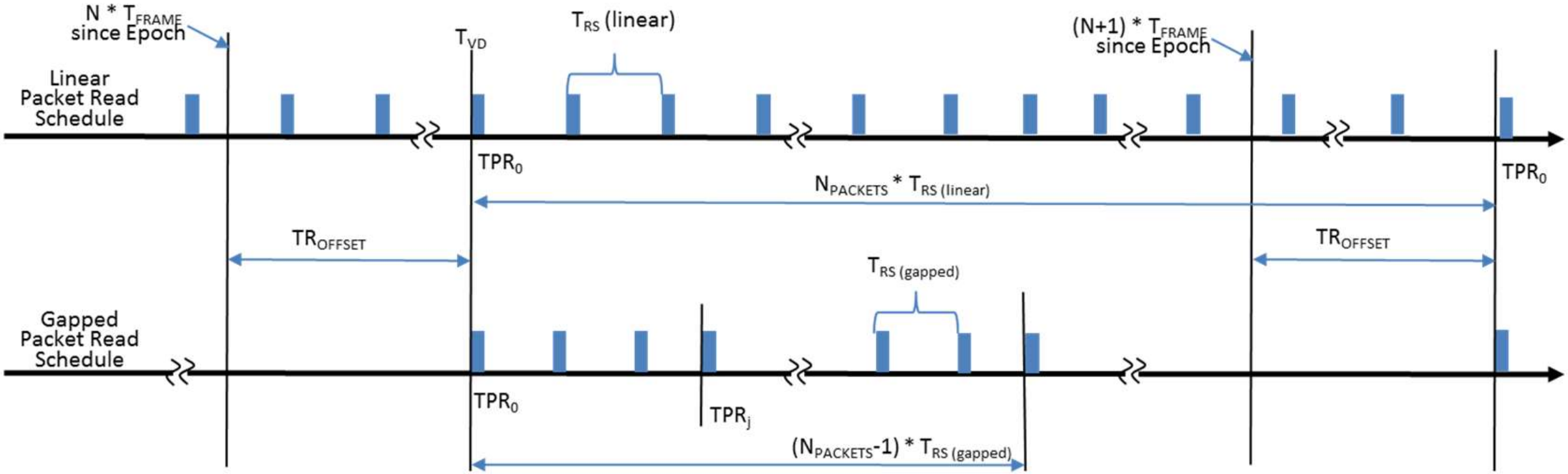


Two Constraints for ST 2110-20 Senders

- Network Compatibility Model
  - Ensures streams will not overflow buffers inside network devices
  - Scaling factor β of 1.1 means buffers drain 10% faster than they fill
- Virtual Receiver Buffer Model
  - Buffer is modeled as input of every receiver device
    - Note: Must be included in end-to-end system delay
  - Packets read from buffer perfectly, based on video format
  - Buffer not allowed to overflow or underflow
- All senders must comply with both models

IP SHOWCASE™ THEATER

ST 2110-21 Gapped, Linear Packet Schedules



telecompro.tv
IP Video Training and Consulting

Image Source: SMPTE ST 2110-21 Traffic Shaping and Delivery Timing for Video

ST 2110-21 Sender Types

- Three Sender Types: N = Narrow, NL = Narrow Linear, W = Wide
- Type N is designed for real-time capture and processing (live events)
  - Maximum required receiver buffer is about 9 packets in gapped mode
  - Model assumes TR_{OFFSET} of a couple of video lines from SMPTE Epoch
  - Small buffer means limited delay passing through each device in systems
  - Pixels inside packets “roughly” in sync with pixels in SDI
- Type NL is linear version of N – no gaps corresponding to SDI VANC
- Type W is designed to support software-based video sources (graphics)
  - Maximum receive buffer is 720 packets in some popular formats
  - Larger buffer can handle packet bursts more easily
  - Bursty transmission is more common to software-based senders

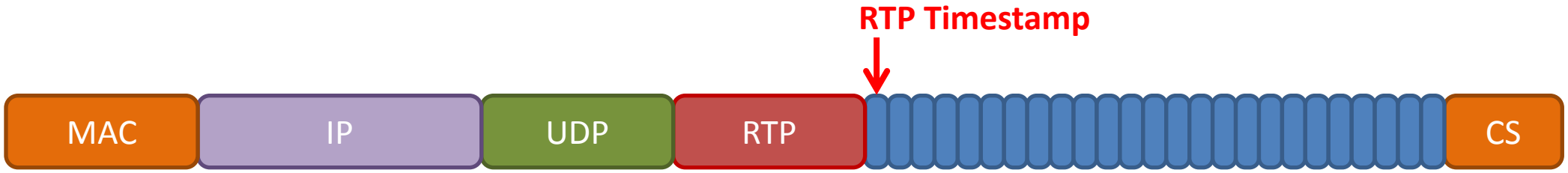
IP SHOWCASE™ THEATER

ST 2110-21 Sender/Receiver Compatibility

Receiver Type	Type N Sender	Type NL Sender	Type W Sender
Type N Synchronous Narrow	Mandatory	Optional	No
Type W Synchronous Wide	Mandatory	Mandatory	Mandatory
Type A Asynchronous	Mandatory	Mandatory	Mandatory

- Synchronous Receivers must have clock locked to Sender
- Synchronous Narrow Receivers are only required to work with Senders that use the default TR_{OFFSET}

ST 2110-30 Audio Encapsulation



- Multiple Audio Samples (16 or 24 bit)
- Grouped into one RTP packet
- Placed into UDP packet
- IP packet header attached
- Wrapped into Ethernet Frame

ST 2110-30 Audio

- Based on AES67
  - 48 kHz, 24-bit linear encoding must be supported in all devices
- Zero Offset Media Clock
  - Forces all media clocks to be tied to common time base
- Audio Channel Grouping
  - How audio channels relate to each other in a stream
- Receiver Classifications
  - Three levels of receiver performance
- Packet size limit 1440 = 1460 – (12 (RTP) + 8 (UDP))
- No need for SIP or other connection management

Importance of "`ptime`"

- Audio streams are divided into fixed duration packets
  - Common size is 1 msec, signaled using "`a=ptime:1`" attribute
- Number of samples from a channel depends on sampling rate
  - For example, 48 kHz has 48 samples in 1 msec
  - Each sample could be 2 bytes (16 bit audio) or 3 bytes (24 bit audio)
  - Thus, 1 msec of 48 kHz, 24-bit audio is 48 * 3 = 144 bytes
- Number of channels in a packet limited by payload size
  - Total RTP audio payload is 1440 bytes
  - Jumbo frames not allowed for audio

IP SHOWCASE™ THEATER

ST 2110-30 Receiver Classifications

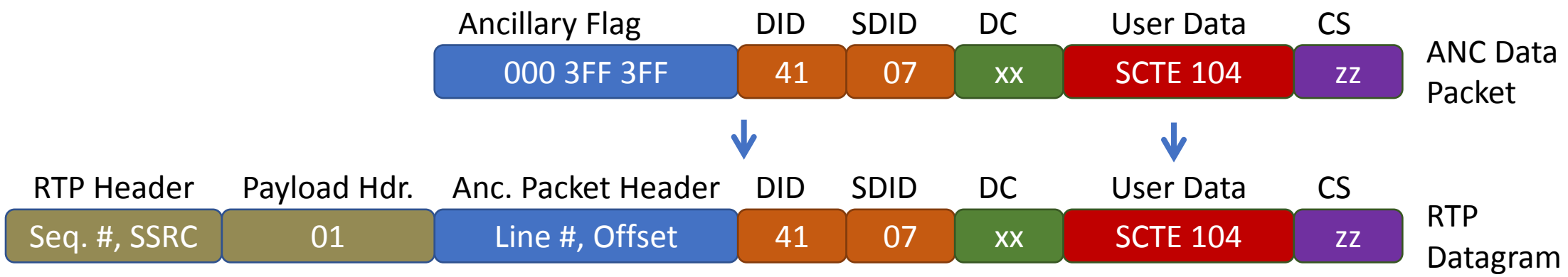
Required Sampling Rates and Packet Times	A	AX	B	BX	C	CX
48 KHz, 1 msec	8	8	8	8	8	8
48 KHz, 125 µsec			8	8	64	64
96 KHz, 1 msec		4		4		4
96 KHz, 125 µsec				8		32

ST 2110-30 Audio Channel Grouping Symbols

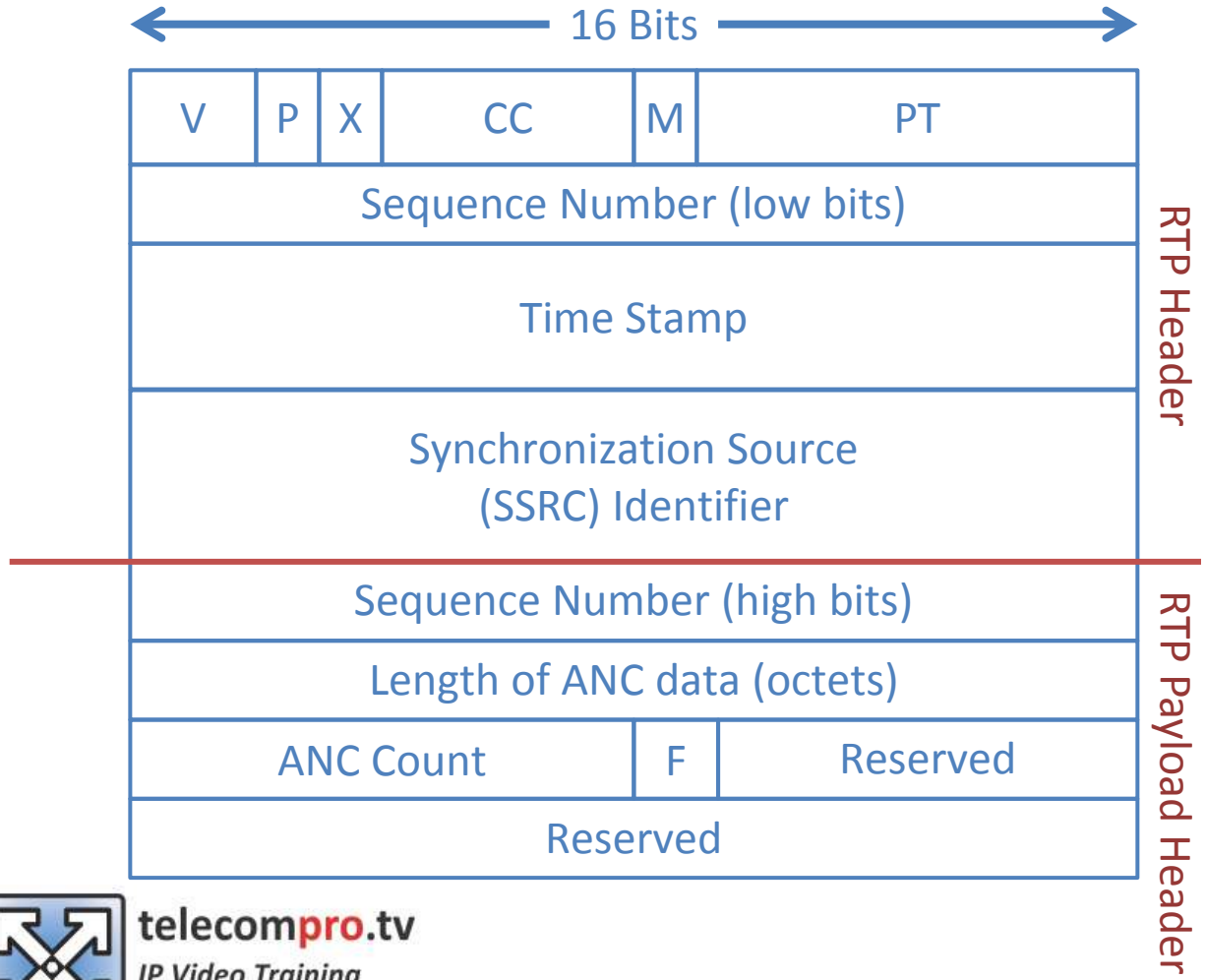
Channel Grouping Symbol	Quantity of Audio Channels in group	Description of group	Order of Audio Channels in group
M	1	Mono	Mono
DM	2	Dual Mono	M1, M2
ST	2	Standard Stereo	Left, Right
LtRt	2	Matrix Stereo	Left Total, Right Total
51	6	5.1 Surround	L, R, C, LFE, Ls, Rs
71	8	7.1 Surround	L, R, C, LFE, Lss, Rss, Lrs, Rrs
222	24	22.2 Surround	Per SMPTE ST 2036-2, Table 1
U01…U64	Unn where nn is the number of channels in group	Undefined	Undefined

2110-40 Ancillary Data

- Extract ancillary data packets from VANC or HANC
  - Captions, time code, ad triggers, etc.
  - Place them into RTP packets with custom header
- Line numbers are based on SDI line numbering
  - Don't match 2110-20 line numbers

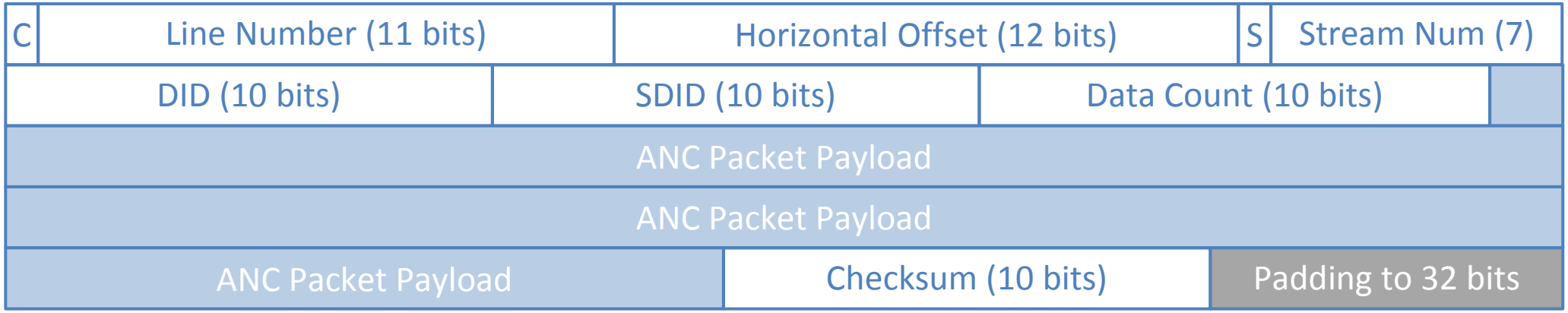


ST 2110-40 Ancillary Packet Payload Header



- 32-bit Sequence Number (same as ST 2110-20 video)
- Length of ANC data = Number of octets of all ANC packet headers, ANC payloads, and stuffing
- ANC Count = Number of ANC packets in this payload
- F = Field flag indicates source of ANC packets in this RTP packet, as follows:
  - 00 = Progressive video frame or no source specified
  - 01 = Not valid
  - 10 = First field of an interlaced or PsF frame
  - 11 = Second field of an interlaced or PsF frame

ST 2110-40 ANC Packet Format



- Each ANC packet in the RTP payload has its own header
- Color channel flag: C=1 – ANC packet is from HD color difference channel. C=0 in all other cases
- Line Number and Horizontal Offset refer to SDI raster values
- S=1 Multiple streams comprise the format of the original video signal containing the ANC packets
- Stream number indicates where the ANC packets were located within a multi-stream signal
- DID, SDID, Data Count, Packet Payload and Checksum are exact 10-bit values from ANC packet
- For each ANC packet within the RTP payload, padding makes the total number of bits a multiple of 32

Lossless Compression

- Visually lossless compression cannot be seen by observer
  - Some data must always be removed
  - Done so as to be invisible to human viewer
  - Can have very low latency – using slice-based compression
- Popular codecs available
  - VC-2 DIRAC from BBC – RFC 8450
  - Also JPEG XS – draft-lugan-payload-rtp-jpegxs-01
- 2:1 to 8:1 compression ratios
  - 3Gbit/s SDI compressed to 1.5 to 0.5 Gbit/s

Forthcoming: SMPTE ST 2110-22

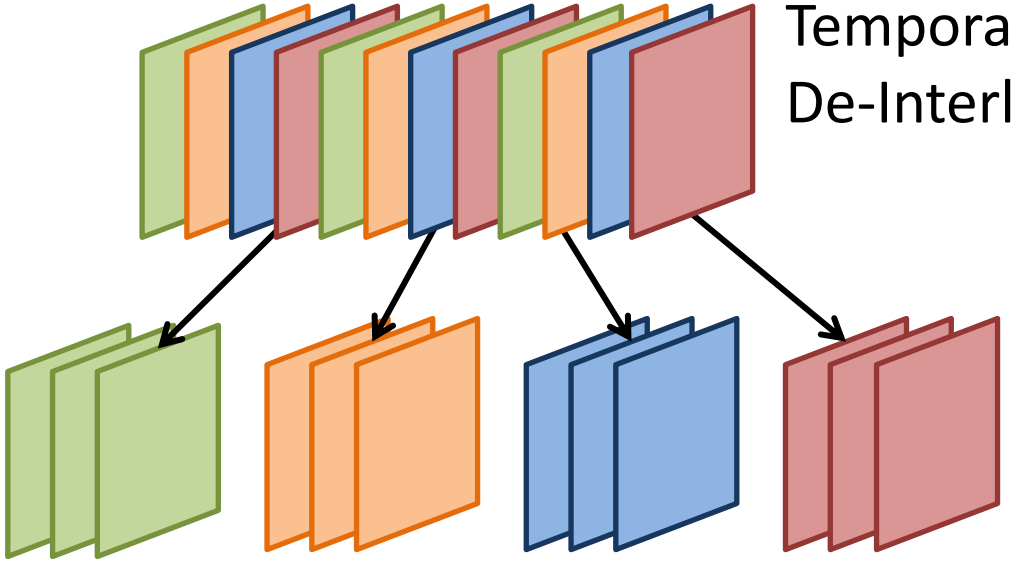
- Current Title: “Professional Media over Managed IP Networks: Constant Bit-Rate Compressed Video”
  - Supports CBR compression formats such as VC2
  - Must be a registered RTP media type as per RFC 4855
  - RTP Clock rate of 90 kHz
  - Must conform to either “NL” or “W” network compatibility model of ST 2110-21; virtual receiver buffer model does not apply

Forthcoming: RP 2110-23

- Working Title: “Single Video Essence Transport over Multiple ST 2110-20 Streams”
- Idea is to have a system where multiple low-bandwidth streams can be used to transport one high-bandwidth signal
  - High resolution streams, such as UHD1/4K or UHD2/8K
  - High frame rate streams, such as those over 100 fps
  - Also known as “multiport”
- Each sub-stream is a valid ST 2110-20/2110-21 stream
  - Timestamps tied to original frames
  - Comply with timing models

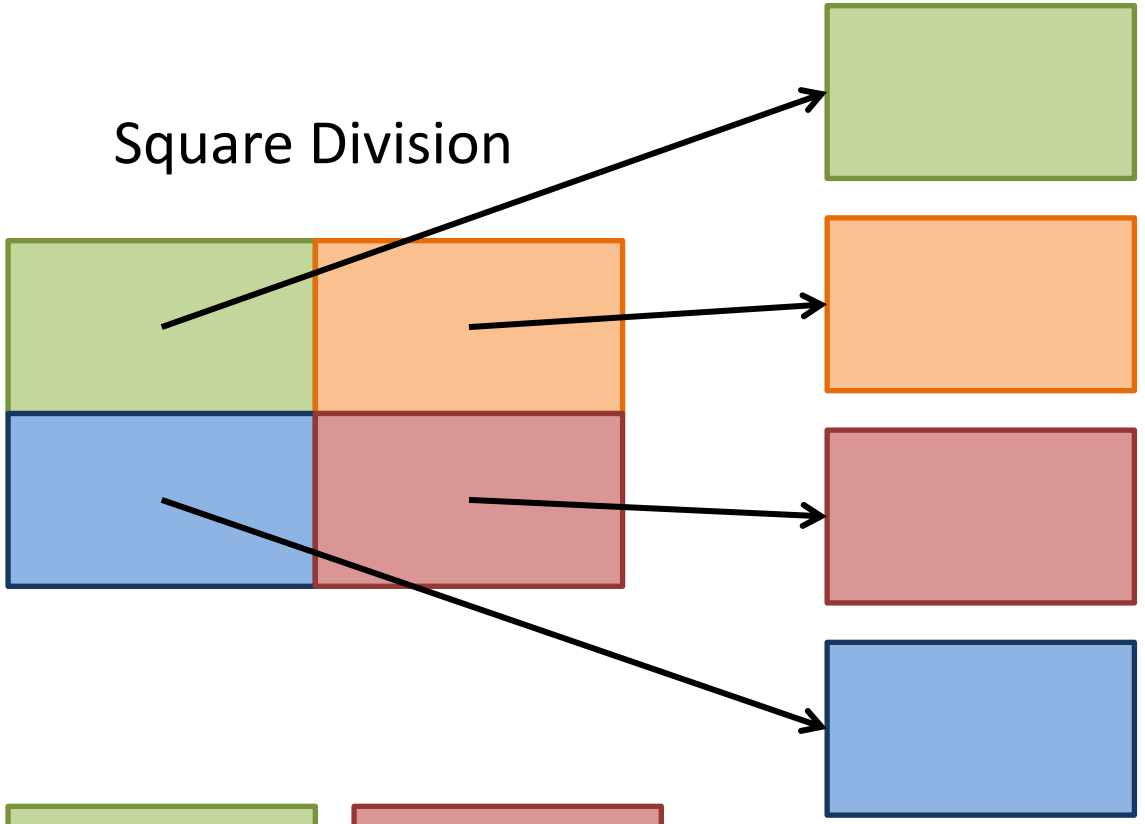
IP SHOWCASE™ THEATER

Three Methods to Split Stream

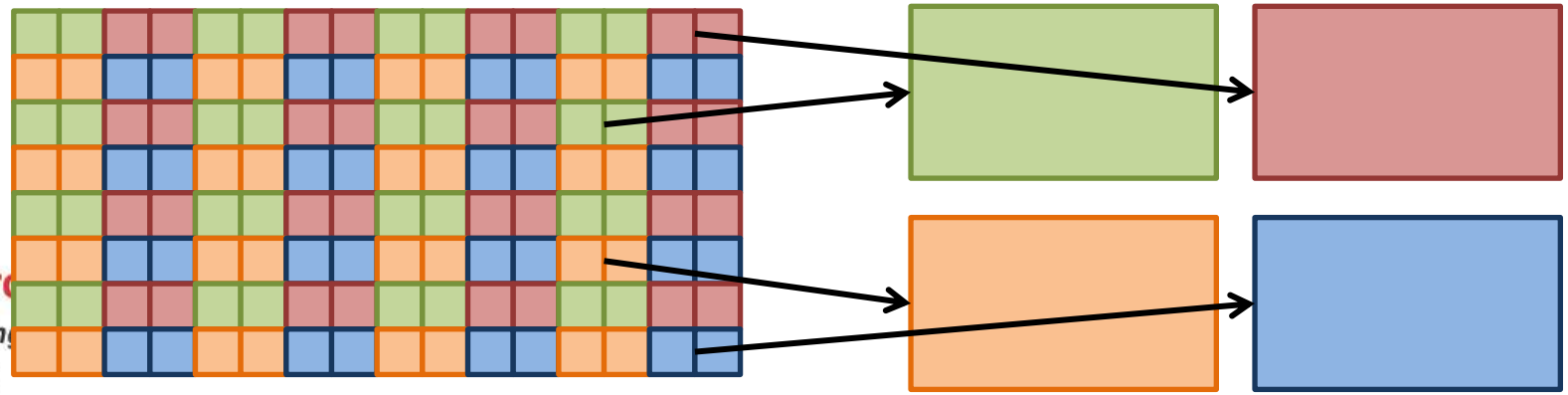


Temporal
De-Interleave

Square Division



Two-Sample
Interleave



RTP Timestamps

- RTP Timestamps depend on rate of individual Media Clock
  - Video: 90 kHz
  - Audio: 48 kHz or 96 kHz
- RTP Timestamps for 11:00:00 am PDT Apr. 8, 2019 (UTC -7:00)
  - Seconds since SMPTE/PTP Epoch: n = (17,995.75 days * 86,400 + 37)
  - Video Timestamp = mod 2^{32} (n*90,000) = 625,859,024
  - 48 kHz Audio Timestamp = mod 2^{32} (n*48,000) = 2,624,440,704
  - 96 kHz Audio Timestamp = mod 2^{32} (n*96,000) = 953,914,112

Timestamp Rollover Times

- RTP Timestamp field is 32 bits
  - Therefore, timestamps will rollover every 2^{32} clock ticks
  - Rollover time == time between points when timestamp is zero
- Each PTP clock frequency will have a different rollover time
  - How many hours for 2^{32} clock ticks?

Clock Frequency	Rollover Time
90 kHz video	13.256 hours
48 kHz audio	24.855 hours
96 kHz audio	12.428 hours

SDP for ST 2110-22

- Format parameters (**`a=fmtp`**) statement must include
  - Image height in lines
  - Image width in pixels
  - TP of either 2110TPNL or 2110TPW
  - Optional value of CMAX if different from default
- Bit rate parameter "**`b=AS:<bandwidth>`**" must be included
  - Bandwidth is in kilobits/second calculated over one frame period
- SDP must include a frame rate statement, either
  - **`a=framerate xx.yy`** (as a decimal number)
  - **`exactframerate=M/N`** (as a ratio of two integers) in "fmtp"

CURATED BY

VIDEO SERVICES FORUM

wes.simpson@gmail.com

www.telecompro.tv