

# SMPTE ENGINEERING GUIDELINE

## Image Identification, Alignment, Transport and System Guidance for Stereoscopic (S3D) or Multi-Camera Array



<b>Table of Contents</b>	<b>Page</b>
Foreword .....	2
Intellectual Property .....	2
Introduction.....	2
1 Scope .....	3
2 Conformance Notation .....	3
3 Terms and Definitions .....	3
4 Overview .....	4
5 Image Identification.....	5
6 Image Alignment.....	6
7 Transport.....	6
7.1 Network.....	6
7.2 SDI.....	6
7.3 Payload ID .....	8
8 System Guidance.....	9
8.1 Time Stamp .....	9
8.2 Genlock.....	9
8.3 Processing.....	9
8.4 Disparity Maps and Depth Maps .....	11
8.5 Record/Playback Device Expectations.....	11
Bibliography (Informative) .....	12

## Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in its Standards Operations Manual.

SMPTE EG 2076-2 was prepared by Technology Committee 32NF.

## Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE must not be held responsible for identifying any or all such patent rights.

## Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

Stereoscopic (S3D) or Multi-Camera Array imaging systems deliver two or more images to a downstream process that can be used for stereoscopic display, light field computation or a Multi-View display.

To maintain the fidelity of S3D or Multi-Camera Array images it is critically important that the spatial and temporal alignment of the images be maintained at all times. Loss of alignment between the images will result in the degradation or complete destruction of the stereoscopic or Multi-Camera Array for post processing and ultimately the viewer.

This document highlights the methods and specific issues to be aware of when dealing with S3D or Multi-Camera Array images within a system. This document does not address audio or control systems as they relate to S3D or Multi-Camera Array images.

## 1 Scope

This Engineering Guideline provides an overview of the identification, synchronization and transport of Stereoscopic (S3D) or Multi-Camera Array motion picture and television images over SDI transport systems.

## 2 Conformance Notation

This Engineering Guideline is purely informative and meant to provide tutorial information to the industry. It does not impose Conformance Requirements and avoids the use of Conformance Notation.

Engineering Guidelines frequently provide tutorial information about a Standard or Recommended Practice and when this is the case, the user ought to rely on the Standards and Recommended Practices referenced for interoperability information.

## 3 Terms and Definitions

For the purposes of this document, the following terms and definitions apply.

### 3.1 left eye Le

Abbreviation for Left Eye.

### 3.2 right eye Re

Abbreviation for Right Eye.

### 3.3 three dimensional 3D

Acronym for Three Dimensional. Adding depth as the third visual dimension.

### 3.4 stereoscopic 3D S3D

Acronym for Stereoscopic 3D.

### 3.5 S3D Image Content

Stereoscopic Image content, which results in the viewing of a scene with the perception of depth to the observer.

### 3.6 Stereoscopic

Relating to the use of binocular vision to create the perception of depth in an image.

### 3.7 Multi-Camera Array

Two or more cameras aligned to capture a single scene coincident in time for downstream light field image processing.

### 3.8 Coincident in Time

With respect to dual image signals for stereoscopic television, this means that not only are the two image signals "genlocked", but that they represent the same moments in time for the image displayed.

### 3.9 Camera

As defined here for this document includes the lens and all associated equipment required to create digital representations of moving images. The camera is generally capable of stand-alone operation.

### 3.10 Camera Systems

As defined here for this document, a *camera system* consists of all *cameras* and their associated equipment to create S3D content or for computational light field image processing.

### 3.11 Genlock

Abbreviation of “[sync] Generator Lock.” Genlock is a technique for locking a device’s internal sync structures (and thus image structures) to a common external reference (a “sync generator”).

### 3.12 Temporal Alignment

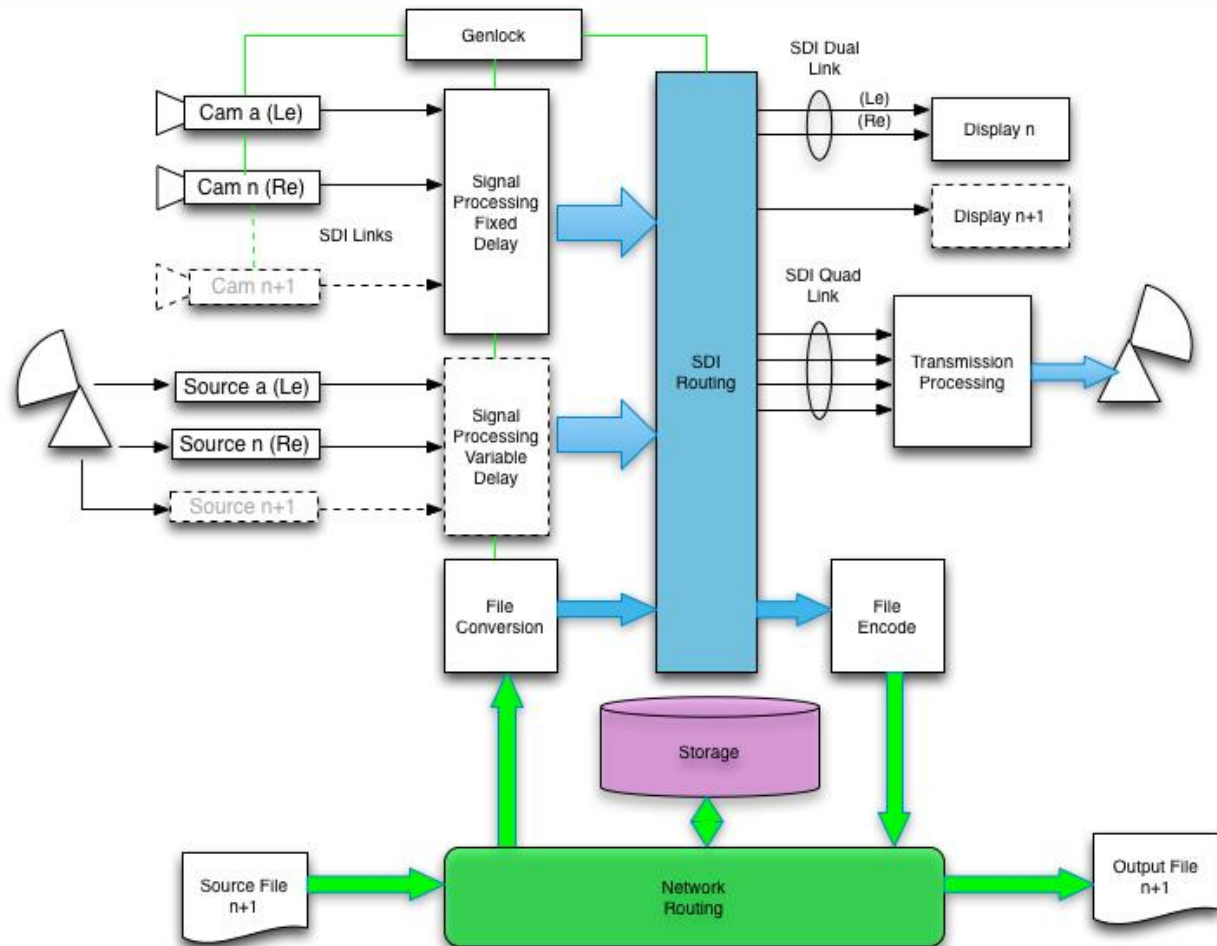
With respect to the capture mechanisms of digital motion picture and television cameras, the capture sensor systems need to be temporally synchronized in order to maintain proper representation of motion when using more than one sensor or device to capture S3D or Multi-Camera Array images for presentation.

## 4 Overview

S3D or Multi-Camera Array images must maintain their spatial and temporal alignment so that downstream devices can process and display the content correctly to the viewer. Stereoscopic content consists of a pair of images normally designated Left Eye (Le) and Right Eye (Re) that must remain coincident in time. Multi-Camera Array content consists of two or more images that must remain coincident in time.

This document is mostly centered around Live Event systems where real-time processing and transmission are a basic requirement for the system. (i.e. Broadcast) Non real-time workflows, such as file based processes, are mentioned briefly but are not extensively covered here. If proper identification of frames is maintained for file-based systems, then there are fewer opportunities for synchronization issues to occur.

Below for reference is an example figure showing the key components of a Live Event system. This is a general example and does not show a complete system nor does it attempt to show each process that could be used. Instead it attempts to illustrate the key processes where synchronization can be affected.



**Figure 1 – Example Live Event System**

Identification of the images becomes a critical element for maintaining alignment throughout the system. Without knowing the relationship between the images, designing systems to maintain alignment will be difficult if not impossible. In the following sections of this Engineering Guideline it steps through the key issues for Identification, Alignment and the Transport of S3D or Multi-Camera Array images.

## 5 Image Identification

During Image acquisition or creation, an image pair (in the case of S3D) ought to be identified as Left Eye (Le) and Right Eye (Re). Images that are part of a Multi-Camera Array ought to be identified with a non-duplicate identifier depending on the intent of the image captured or created coincident in time. For example Camera: 1,2,3... a, b, c... Top Left, Top Center, Top Right...

Along with identifying the image “intent”, the system ought to provide temporal identification such as Time code or frame count.

File based workflows have many ways of tagging image frames for identification. We will not explore these as there are too many methods and this is not the core objective of this document. One ought to refer to the specification of the file format that is used and decide on the best method in which to tag each image frame.

This document will provide methods in the sections below to tag the image frames in SDI transport streams.

## 6 Image Alignment

Each individual camera image derived from a S3D or Multi-Camera Array system must maintain temporal alignment. This alignment must the time of capture of moving images.be maintained throughout the camera system, without change or drift during

This alignment must be maintained throughout the production system as well, without change or drift during the time of capture or generation of the images. This alignment must be maintained and preserved to a recording device, a display device and to the output of the system.

For optimum results there ought to be no fixed offsets or temporal delays between each of the images. If this is unavoidable, a fixed offset that does not change over time can be used provided it is transmitted and recognized by the downstream systems required to process the images.

## 7 Transport

When transporting S3D on all interfaces it is extremely important to be able to identify the Left Eye image stream from the Right Eye stream. If these image streams are not identified, or are improperly identified, then it is difficult to reliably connect the links to the correct inputs and outputs of equipment. In the case where there is no identification applied at the source, it is very important to insert the payload ID as early as possible in the production system. There are devices available from some equipment vendors that allow insertion of the correct payload ID as well as swapping the signals if they are incorrectly identified.

### 7.1 Network

More and more common IT network systems are being used to transport images these days. The bulk of this is used for non-real-time transport. There are some cases where these networks are used for real-time transport. In either case this document will not address this as there are many different types of file formats and network protocols. Proper identification is key when using these systems and ought to be maintained at all times.

### 7.2 SDI

For many years SMPTE ST 292-1 has been widely used to deploy HDTV systems. It defines the bit-serial data structure and the coaxial cable interface specifications for a nominal 1.5 Gb/s Signal/Data Serial Interface to carry either 4:2:2 10 bit 1280x720, 1920x1080 or 2048x1080 active pixel formats defined by SMPTE ST 296, SMPTE ST 274 and SMPTE ST 2048-2 and which can be mapped into a 1.5 Gb/s payload. SMPTE ST 292-1 also maps audio and other ancillary data into the 1.5 Gb/s payload.

#### 7.2.1 4:2:2 Sampling

With the advent of S3D production, it became necessary to carry two of these image streams in such a way as to maintain the spatial and temporal relationships between the Left Eye stream and the Right Eye stream. SMPTE ST 292-2 was developed to standardize a Dual Link 1.5 Gb/s (nominal) interface for S3D. That interface uses a standard SMPTE ST 292-1 interface for each of the Left and Right Eye streams and also defines a payload identifier (see Section 7.3 below) that will identify the Left Eye (Le) or Right Eye (Re) images, audio and other associated ancillary data.

SMPTE ST 425-1 defines the data structures for the carriage of SMPTE ST 296, SMPTE ST 274 and SMPTE ST 2048-2 images that exceed the 1.5 Gb/s capacity of SMPTE ST 292-1 and permits Level A and Level B Dual Link mappings of these images onto these data structures. In addition, SMPTE ST 425-1 provides a Level B Dual stream mapping which maps two 1.5 Gb/s payload images onto these data structures. SMPTE ST 424 defines the bit serial data structure and coaxial cable specifications for carrying the payloads mapped by SMPTE ST 425-1 serial interface.

Carrying two related images on a Dual Link interface has always presented challenges in keeping the images together. SMPTE ST 425-2 was developed to standardise the carriage of S3D streams carried by Dual Link 1.5 Gb/s SMPTE ST 292-2 interfaces on a single 3 Gb/s interface using these Level B Dual Stream mapping rules. It essentially carries the complete payload of the SMPTE ST 292-2 interface and applies a new payload identifier to indicate the carriage of S3D images on a Single Link 3 Gb/s interface. SMPTE ST 424 still specifies the serial interface for SMPTE ST 425-2 as it utilizes the fundamental data structures defined in SMPTE ST 425-1.

The 4:2:2/10-bit image formats defined by SMPTE ST 274 and SMPTE ST 2048-2 for higher than 30 frames per second frame rates do not fit into a 1.5 Gb/s payload, and can be carried on a SMPTE ST 425-1 3 Gb/s coaxial interface. SMPTE ST 425-4 was developed to standardize a Dual Link nominal 3 Gb/s interface for S3D. This interface uses a standard SMPTE ST 425-1 Level A or Level B Dual Link interface for each of the left and right eye streams and also defines the payload identifier that will identify the Le and Re images, audio and other associated ancillary data. Once again, SMPTE ST 424 specifies the serial interface for SMPTE ST 425-4 as it utilizes the fundamental data structures defined in SMPTE ST 425-1 for each Eye.

**Table 1 – 4:2:2 10 Bit Image Transport Summary**

S3D or Mono	Image Format	Nominal Frame Rate (Frames/sec)	HDSDI Data Mapping	HDSDI Serial Interface	Link Data Rate (nominal Gb/s)	Number of Links	Interface Data Rate (nominal Gb/s)
Mono	1280x720	24 to 60	ST 292-1	ST 292-1	1.5	1	1.5
	1920x1080	24 to 30					
	2048x1080						
S3D	1280x720	24 to 60	ST 292-2	ST 292-1	1.5	2	3
	1920x1080	24 to 30					
	2048x1080						
	1280x720	24 to 60	ST 425-2	ST 424	3	1	3
	1920x1080	24 to 30					
	2048x1080						
Mono	1920x1080	48 to 60	ST 425-1	ST 424	3	1	3
	2048x1080						
S3D	1920x1080	48 to 60	ST 425-4	ST 424	3	2	6
	2048x1080						

### 7.2.2 4:4:4 Sampling

The 4:4:4/10-bit image formats defined by SMPTE ST 274 and ST 2048-2 for up to and including 30 frames per second frame rates do not fit into a 1.5 Gb/s payload, and can be carried on a SMPTE ST 425-1, 3 Gb/s coaxial interface. SMPTE ST 425-4 was developed to standardize a Dual Link nominal 3 Gb/s interface for S3D. This interface uses a standard SMPTE ST 425-1 Level A or Level B Dual Link interface for each of the left and right eye streams and also defines the payload identifier that will identify the Le and Re images, audio and other associated ancillary data. Once again, SMPTE ST 424 specifies the serial interface for SMPTE ST 425-4 as it utilizes the fundamental data structures defined in SMPTE ST 425-1 for each Eye.

The 4:4:4/10-bit image formats defined by SMPTE ST 274 and SMPTE ST 2048-2 for higher than 30 frames per second frame rates do not fit into a 3 Gb/s payload, and can be carried on a SMPTE ST 425-5 Dual Link 3 Gb/s coaxial interface. SMPTE ST 425-6 was developed to standardize a Quad Link nominal 3 Gb/s interface for S3D. This interface uses a standard SMPTE ST 425-5 Level A or Level B Dual Link interface for each of the left and right eye streams and also defines the payload identifier that will identify the Le and Re images, audio and other associated ancillary data. Once again, SMPTE ST 424 specifies the serial interface for SMPTE ST 425-6 as it utilizes the fundamental data structures defined in SMPTE ST 425-1.

**Table 2 – 4:4:4 10-Bit Image Transport Summary**

S3D or Mono	Image Format	Nominal Frame Rate (Frames/sec)	HDS DI Data Mapping	HDS DI Serial Interface	Link Data Rate (nominal Gb/s)	Number of Links	Interface Data Rate (nominal Gb/s)
Mono	1280x720	24 to 60	ST 425-1	ST 424	3	1	3
	1920x1080	24 to 30					
	2048x1080						
S3D	1280x720	24 to 60	ST 425-4	ST 424	3	2	6
	1920x1080	24 to 30					
	2048x1080						
Mono	1920x1080	48 to 60	ST 425-5	ST 424	3	2	6
	2048x1080						
S3D	1920x1080	48 to 60	ST 425-6	ST 424	3	4	12
	2048x1080						

**7.3 Payload ID**

If S3D Imaging Systems serial digital outputs support the SMPTE ST 352 Payload ID, then with the S3D information encoded into the Payload ID you can identify the difference between the Le and the Re.

SMPTE ST 352 defines a payload identifier ancillary data packet that is used to uniquely identify both the payload type and the interface it is travelling on.

For the SMPTE ST 292-1 1.5 Gb/s serial interface, this payload identifier is optional, and in fact most existing implementations do not use the payload identifier. Many early implementations of S3D transport have been done by using two 1.5 Gb/s interfaces, and so it is not possible to identify where the Left Eye image and Right Eye image are.

When SMPTE ST 292-2 was standardized in 2011, it mandated the use of the payload identifier on the S3D interface, in order to identify not only which link carries the left and right image streams, but also to identify the audio carried on the interface. The first byte of this payload ID specifies it is carrying S3D images on a 1.5 Gb/s dual link interface. The second and third bytes of the ID packet specify the image structure and sampling structure, and the fourth byte species which link carried the left and right image streams and whether the audio on the right eye is a duplicate of the left eye audio or is additional audio. SMPTE ST 425-2 interfaces carry all of the payload of the SMPTE ST 292-2 Dual Link interface on one 3 Gb/s interface, so the payload ID used is identical to the SMPTE ST 292-2 payload ID except for byte 1 which specifies it is carrying S3D images on a single link 3 Gb/s interface.

For all of the SMPTE ST 425 3 Gb/s interfaces, the SMPTE ST 352 payload ID is mandatory. For the stereoscopic interfaces (SMPTE ST 425-4 and SMPTE ST 425-6), the payload ID identifies not only which link(s) carry the left and right image streams, but also to identify the audio carried on the interface.

## **8 System Guidance**

### **8.1 Time Stamp**

The camera capture or content generation system ought to have the ability to sequentially time stamp each frame of each image stream recorded to support alignment of frames later on in post-production. This time stamp must conform to SMPTE ST 12-1.

#### **8.1.1 Support of SMPTE ST 12-2 time code**

SDI transport systems ought to include SMPTE ST 12-2 time code on all separate image streams. (ex. Le, Re or individual streams in a Multi-Camera Array transport.)

#### **8.1.2 Support of Two-Frame Marker**

For progressive image formats of 48/(1.001) Hz, 50 Hz and 60/(1.001) Hz, one ought to use the two-frame marker method described in SMPTE ST 2051.

### **8.2 Genlock**

Temporal alignment between the signals of a S3D image pair or Multi-Camera Array can be established using an external genlock reference.

All devices within a facility, which process or manipulate a S3D or Multi-Camera Array content ought to be genlocked to the facility Master Sync Generator. Devices can support at a minimum one of the following genlock signals: SMPTE ST 318, SMPTE ST 274, SMPTE ST 296 or SMPTE ST 2059-1/2. The output timing of such devices can be set identically such that the signals are temporally coincident.

While it is not required that the devices be locked to an external reference (Stand-alone operation), it is required at a minimum so that they can be locked to each other.

### **8.3 Processing**

Where two or more channels containing S3D or Multi-Camera Array images are subjected to processing, it is desirable that the controls for the two or more processors or processing channels be ganged together such that the control operation is performed identically on all channels. Where ganged control is not available, operational care is to be considered to ensure that all (non-ganged) controls are set identically for all channels. When control operations are taking place, it is understood that differential timing between channels can occur.

#### **8.3.1 Delay**

Delay can occur in many parts of the Live Event ecosystem. There are two basic types of delay that can effect S3D or Multi-Camera Array systems, fixed delay and variable delay.

##### **8.3.1.1 Fixed Delay**

Certain processes can take up to one frame or longer to complete in a facility system. These cause delays in the signal. With S3D or Multi-Camera Array content, it is very important to match the delay equally across all channels. Different fixed delays between image streams can disturb the resulting presentation unless these are corrected. This can be done after the fact in a non real-time system by adjusting one stream in relation to

the other to realign the content. It can also be accomplished by inserting delay into the image stream(s) that are advanced to correct the issue for a real-time system.

### **8.3.1.2 Variable Delay**

Variable delay, in particular, frame synchronizers with 'smart' drop / repeat algorithms might exhibit different drop / repeat characteristics due to differences in input timing, despite having their outputs aligned. Care must be taken to ensure that such devices drop and repeat frames identically to all image channels under all expected operational timing conditions. Care can be taken to ensure that the behavior of the image channels are identical at all times for all cases of expected input to output timing relationships.

Frame synchronization of a S3D or Multi-Camera Array images require that the synchronizers drop / repeat frames at the same time as each other, and thus might need to be designed for this purpose. Alignment might be accomplished by locking the genlock reference of each frame synchronizer to the same reference source and setting the output timing of each so they are coincident at the output. This might be further augmented by an interconnection between synchronizers which ensures that the dropping / repeating of frames is done in tandem.

### **8.3.2 Scaling**

Where scaling, aspect-ratio conversion or other spatial modification of the images is performed, the positioning and geometric scaling of the resultant images must be identical for each image, resulting in the maintenance of the spatial alignment between the images.

Take care when applying spatial modification to Multi-Camera Array content especially if the intent is for light field processing. Incorrect spatial modification can destroy the ability to perform meaningful computation on the images.

When DVE (Digital Video Effects) are employed, the transform of the DVE must be identical for each image on a frame-by-frame basis. This applies equally to the beginning and end frames of the DVE move as well as for all frames within and during the move.

When deinterlacing is required for processing, the deinterlacers for each image must operate in the same modes and process the images using identical processes. The temporal delay of each deinterlacer must be the same. Where adaptive (based on the motion in the images) deinterlacing is used, the deinterlacers must make the same mode change decisions at the same time. This ought to be further augmented by interconnection between the deinterlacers which ensures that mode changes are synchronized.

When AFD / bar data is used to convey image aspect ratio, the AFD / bar values must be the same for each image at all times.

Equipment which receives AFD / bar data and uses it to alter the image must perform the same transformation on each image at the same instant in time. The resultant AFD / bar data must be the same for each image and bound to each image at the same instant in time.

### **8.3.3 Color Correction**

When color correcting S3D or Multi-Camera Array images, extreme care ought to be taken to balance the two or more images equally. Large differences, especially in luminance, will have a degrading affect on the image. This is very noticeable in Stereoscopic fusing of images. It can also effect downstream processes in the case of light field computations.

#### **8.4 Disparity Maps and Depth Maps**

Disparity and Depth maps can be extremely useful for S3D and Light Field processing (segmentation, transmission codecs) and for display purposes. One such type is described in SMPTE ST 2087.

Disparity maps and depth maps need to be carefully preserved and maintained in sync with the image streams they are intended to be used with.

#### **8.5 Record/Playback Device Expectations**

Where image channels containing a S3D or Multi-Camera Array content are recorded to storage media, channels can be recorded simultaneously with the same time code or other temporal marker such that at playback there is no ambiguity as to the association of the images.

Where image channels containing a S3D or Multi-Camera Array content are played back from stored media, the image channels must output the correct images coincident in time at all times while stopped, cueing, jogging or playing, including off-speed play, and the time code or other temporal markers on each channel must contain identical values for each set of images.

## Bibliography (Informative)

SMPTE ST 12-1:2014, Time and Control Code

SMPTE ST 12-2:2014, Transmission of Time Code in the Ancillary Data Space

SMPTE ST 274:2008, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE ST 292-1:2012, 1.5 Gb/s Signal/Data Serial Interface

SMPTE ST 292-2:2011, Dual 1.5 Gb/s Serial Digital Interface for Stereoscopic Image Transport

SMPTE ST 296:2012, 1280 x 720 Progressive Image 4:2:2 and 4:4:4 Sample Structure — Analog and Digital Representation and Analog Interface

SMPTE ST 318:2015, Television and Audio — Synchronization of 59.94 or 50 Hz Related Video and Audio Systems in Analog and Digital Areas — Reference Signals.

SMPTE ST 352:2013, Payload Identification Codes for Serial Digital Interfaces

SMPTE ST 424:2012, 3 Gb/s Signal/Data Serial Interface

SMPTE ST 425-1:2011, Source Image Format and Ancillary Data Mapping for the 3 Gb/s Serial Interface

SMPTE ST 425-2:2012, Source Image Format and Ancillary Data Mapping for Stereoscopic Image Formats on a Single-Link 3 Gb/s Serial Interface

SMPTE ST 425-4:2012, Dual 3 Gb/s Serial Digital Interface for Stereoscopic Image Transport

SMPTE ST 425-5:2015, Image Format and Ancillary Data Mapping for the Quad Link 3 Gb/s Serial Interface

SMPTE ST 425-6:2014, Quad 3 Gb/s Serial Digital Interface for Stereoscopic Image Transport

SMPTE ST 2048-2:2011, 2048 x 1080 Digital Cinematography Production Image FS/709 Formatting for Serial Digital Interface

SMPTE ST 2051:2014, Two-Frame Marker for 48/(1.001) Hz, 50 Hz and 60/(1.001) Hz Progressive Digital Video Signals on 1.5 Gb/s and 3 Gb/s Interfaces

SMPTE ST 2059-1:2015, Generation and Alignment of Interface Signals to the SMPTE Epoch

SMPTE ST 2059-2:2015, SMPTE Profile for Use of IEEE-1588 Precision Time Protocol in Professional Broadcast Applications

SMPTE ST 2087:2016, Depth Map Representation