

UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE PATENT TRIAL AND APPEAL BOARD

Samsung Electronics Co. Ltd. and Samsung Electronics America, Inc.
"Samsung"
Petitioners

v.

Advanced Coding Technologies, LLC
"ACT"
Patent Owner

INTER PARTES REVIEW OF U.S. PATENT NO. 9,986,303

DECLARATION OF DR. CLIFFORD READER

I declare that all statements made in this declaration on my own knowledge are true and that all statements made on information and belief are believed to be true, and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Date: 1/5/24

By: 

Clifford Reader, Ph.D.

Contents

I.	Professional Background.....	7
A.	Summary.....	7
B.	Education	8
C.	Work Experience	8
D.	Standardization.....	9
E.	Intellectual Property Rights.....	11
F.	Curriculum Vitae.....	12
II.	Digital Video Technologies.....	12
A.	Analog Video.....	13
B.	Digital Video.....	13
C.	Video Coding.....	16
D.	Video Coding Standards	19
1.	Background	19
2.	The MPEG1 Standards	21
3.	The MPEG2 Standard.....	26
4.	MPEG Scalability	27
E.	The H.263 Standard.....	32
F.	The H.264 Standard.....	39
III.	Relevant Legal Standards.....	48
IV.	Person of Ordinary Skill in the Art.....	50
V.	Overview of the Subject Patent US9986303	51
A.	Overview	52

1.	Field of Art.....	52
2.	Background - Problem statement, Admitted Prior Art.....	52
3.	Summary of the alleged '303 Invention.....	52
4.	Brief Description.....	53
5.	Comments	58
B.	Claims	62
C.	Prosecution History of the '303 Patent	65
D.	Comments	67
VI.	Summary of '303 Prior Art.....	68
A.	Overview of Demircin	69
B.	Overview of Kimoto	74
VII.	Claim Construction	77
VIII.	Ground 1 – Claims 1-2 are Obvious under 35 U.S.C. § 103 based on Demircin	78
A.	Analysis.....	78
1.	Claim 1	78
2.	Claim 2	115
IX.	Ground 2 – Claim 1 Is Obvious Under 35 U.S.C. § 103 based on Kimoto in Combination with Demircin	120
A.	Combination of Kimoto and Demircin	120
B.	Reasons to Combine Kimoto and Demircin.....	121
C.	Analysis.....	129
1.	Claim 1	129
X.	Additional Remarks	151

I, Clifford Reader, Ph.D., do hereby declare:

1. I am making this declaration at the request of Samsung in the matter of *INTER PARTES* REVIEW OF U.S. PATENT NO. 9,986,303, “the ’303 Patent.”

2. I am being compensated for my work in this matter at my standard hourly rate of \$750 for consulting services. My compensation in no way depends on the outcome of this proceeding or the content of my testimony.

3. In preparing this Declaration, I considered the following materials:

- (a) U.S. Patent No. 9,986,303 (“the ’303 Patent”) (Exhibit SAMSUNG-1001)
- (b) The Prosecution History of the ’303 Patent (Exhibit SAMSUNG-1002)
- (c) U.S. Patent Publication No. 2008/0181302 (“Demircin”) (Exhibit SAMSUNG-1005)
- (d) U.S. Patent Publication No. 2007/0153916 (“Demircin916”) (Exhibit SAMSUNG-1006)
- (e) U.S. Patent Publication No. 2011/0211631 (“Kimoto”) (Exhibit SAMSUNG-1007)
- (f) The Prosecution History of Kimoto (“Kimoto File History”) (Exhibit SAMSUNG-1008)
- (g) ITU-T Recommendation H.264 (11/07) (Exhibit SAMSUNG-1009)

- (h) U.S. Patent Publication No. 2004/0196972 (“Zhu”) (Exhibit SAMSUNG-1010)
- (i) U.S. Patent No. 8,335,252 (“Chen”) (Exhibit SAMSUNG-1011)
- (j) U.S. Patent No. 8,175,154 (“Luthra”) (Exhibit SAMSUNG-1012)
- (k) U.S. Patent No. 9,094,641 (“Ansley”) (Exhibit SAMSUNG-1013)
- (l) Wu Jiyan, Shang Yanlei, Cheng Bo, Wu Budan, and Junliang Chen; *Loss Tolerant Bandwidth Aggregation for Multihomed Video Streaming over Heterogeneous Wireless Networks*, WIRELESS PERSONAL COMMUNICATIONS: AN INTERNATIONAL JOURNAL 75. 1265-1282. 10.1007/s11277-013-1422-3; Published online October 5, 2013 (“Wu”) (Exhibit SAMSUNG-1014)
- (m) ISO/IEC 11172-2: 1993 (E) (MPEG1 Video) (Exhibit SAMSUNG-1015)
- (n) ISO/IEC 13818-2: 2000 (E) (MPEG2/H.262 Video) (Exhibit SAMSUNG-1016)
- (o) Infringement Contentions, Advanced Coding Technologies LLC v. Samsung Electronics Co., Ltd. et al., 2:22-cv-00499-JRG, E.D. Tex., filed December 30, 2022 (Exhibit SAMSUNG-1017)
- (p) Schwarz et al., *Overview of the Scalable Video Coding Extension of the H.264/AVC Standard*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, v.17, n.9 (September 2007) (Exhibit SAMSUNG-1018)

- (q) ITU-T Recommendation H.262 (02/98) (Exhibit SAMSUNG-1019)
- (r) ITU-T Recommendation H.263 (02/98) (Exhibit SAMSUNG-1020)
- (s) Reader, *Intraframe and Interframe Adaptive Transform Coding*, EFFICIENT TRANSMISSION OF PICTORIAL INFORMATION, SPIE v.66, 108 (1975) (Exhibit SAMSUNG-1021)
- (t) Reader, *MPEG4: Coding for Content, Interactivity, and Universal Accessibility*, OPTICAL ENGINEERING 35(1), 104 (January 1996) (Exhibit SAMSUNG-1022)
- (u) Reader, *MPEG Patents*, in MPEG VIDEO COMPRESSION STANDARD, Chapter 16, at 357-362 (Mitchell et al ed.) (1996) (Exhibit SAMSUNG-1023)
- (v) Radha et al., *The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming Over IP*, IEEE TRANSACTIONS ON MULTIMEDIA, v.3, 53 (March 2001) (Exhibit SAMSUNG-1024)
- (w) H.264 / MPEG-4 Part 10 White Paper, January 31, 2003 (downloaded from <https://github.com/yistLin/H264-Encoder/blob/1cdaf090b63642932ed726b102ed2cf80908edbb/doc/H.264%20%3A%20MPEG-4%20Part%2010%20White%20Paper.pdf> on Jan. 5, 2024) (Exhibit SAMSUNG-1025)
- (x) Karczewicz et al., *The SP- and SI-Frames Design for H.264/AVC*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, v.13 (July 2003) (Exhibit SAMSUNG-1026)

- (y) ISO/IEC 14496-2: 2004 (MPEG4) (Exhibit SAMSUNG-1027)
- (z) ISO/IEC 13818-2: 1996 (MPEG2) (Exhibit SAMSUNG-1028)
- (aa) U.S. Patent No. 6,292,512 to Radha et al. (Exhibit SAMSUNG-1031)
- (bb) U.S. Patent No. 6,697,426 to van der Schaar et al. (Exhibit SAMSUNG-1032)
- (cc) U.S. Patent No. 5,886,736 to Chen (Exhibit SAMSUNG-1033)
- (dd) Apostolopoulos et al., *Video Streaming: Concepts, Algorithms, and Systems*, HP Laboratories (Sept. 18, 2002) (Exhibit SAMSUNG-1034)
- (ee) U.S. Serial No. 11/479,077 to Demircin, Application As Filed for Demircin916 (“Demircin077”) (Exhibit SAMSUNG-1035)

I. Professional Background

A. Summary

4. I am a digital media consultant providing technical, business development and intellectual property consulting services in the areas of digital media including digital imaging, digital video, digital audio and digital speech. Applications include consumer audio and video transmission and storage; video, audio and speech compression; real-time video processing and display; digital speech processing; image/video/audio systems architecture; and image/video/audio chip architecture. I have held this position since 2001 and have consulted for over 80 clients.

B. Education

5. I received my Doctoral degree in 1974 from the University of Sussex, England. My thesis was titled “Orthogonal Transform Coding of Still and Moving Pictures.” The research for my thesis was performed in residence at the Image Processing Institute, University of Southern California, Los Angeles.

6. I received my B. Eng. Degree with Honors in 1970 from the University of Liverpool, England, in the field of electronics.

C. Work Experience

7. From 1970 to 1973 I performed my graduate research in video compression. I was one of the first to perform a type of image coding (adaptive block transform coding) and the first to apply this type of coding to video. This is described in my thesis and summarized in an SPIE paper. SAMSUNG-1021. These techniques underlie the audiovisual coding standards known as MPEG (Moving Picture Experts Group), H.26x, and virtually all other video compression schemes today.

8. From 1975 to 1981 I studied, designed and developed systems for military imaging systems including real-time image and video reconnaissance systems and battlefield management systems.

9. In the early 1980s I taught classes at Santa Clara University, California in digital signal processing and digital image processing.

10. From 1982 to 1989 I architected and led hardware and software engineering teams in the design of systems for real-time imaging for military, medical and earth resources applications.

11. From 1990 to 2001 I architected and led hardware and software engineering teams in the design of semiconductor chips and systems for real-time imaging in digital consumer audio/video applications, including videoconferencing (with speech coding), broadcast TV and DVD.

D. Standardization

12. In 1990 I became an accredited delegate to the Moving Picture Experts Group – MPEG. I, and my team contributed to the technical work for all three parts of the standard – Systems, Video and Audio – and I participated in the management of the MPEG committee.

13. I was Head of Delegation (HoD) to MPEG for the United States in 1991-1992, during which time the MPEG1 standard was completed and the MPEG2 standard was successfully positioned as the global standard for digital video broadcasting and recording (DVD). I was the Editor in Chief of the MPEG1 standard. I personally reviewed and edited all three parts of the standard in detail, and wrote much of the informative annex for the MPEG1 standard. I participated in the technical and management work for development of the MPEG2 standard.

14. I chaired the implementation subcommittee that analyzed MPEG1 Audio (Levels I & II aka MUSICAM; Level III aka mp3), Dolby AC3 and other proposed audio compression algorithms including legacy algorithms for complexity and cost of implementation.

15. I was a co-founder of the MPEG4 standard and chaired the subcommittee from inception for 2-1/2 years, beginning in 1993, following which I chaired the MPEG4 Requirements subcommittee for a further 2 years. These subcommittees established many of the fundamental principles of the MPEG4 standard, including object-based coding, software-based implementation, and development of the bitstream as a syntactic language. *E.g.*, SAMSUNG-1022. MPEG4 focused on audiovisual low bitrate coding, including low bitrate speech coding.

16. The MPEG4 standard in ISO and the H.263 standard in ITU-T were developed collaboratively, with the H.263 Rapporteur and I synchronizing meetings and establishing profiles in each standard that are precisely compatible.

17. I was instrumental in establishing the work on Advanced Audio Coding (AAC).

18. I initiated the work in Synthetic-Natural Hybrid Coding (SNHC), and personally contributed to the work on compression of 3D graphics in the area of error resilient coding.

19. In the early 2000s, I was an invited expert to the Joint Video Team (JVT) established by ISO and ITU to develop the H.264 standard (also denoted MPEG4 Pt. 10, AVC). I was also invited to become an officer of the China National Standards committee AVS, in which I chair the Intellectual Property Rights Subgroup.

20. I have closely followed the developments of the H.265 and H.266 standards.

E. Intellectual Property Rights

21. In 1993 I was hired by CableLabs to be the technical expert for establishing the MPEG Patent Pool (Now MPEGLA). In the course of creating a list of essential IP to practice the standard, I reviewed approximately 10,000 abstracts and 1,000 patents. This is summarized in a chapter of the MPEG book by Mitchell et al. SAMSUNG-1023.

22. Multiple companies have hire me to assist in developing portfolios of their standards-essential patents.

23. In 2002 I was hired by 10 companies to evaluate the standards essential patent environment for the nascent H.264 standard.

24. In 2003 I was hired to evaluate the standards essential patent environment for the nascent national China AVS standards.

25. In 2013 I was hired to evaluate the standards essential patent environment for the nascent AV1 standard.

26. In 2017 I was hired to evaluate the standards essential patent environment for the H.265/HEVC standard.

27. I am the Co-Director of the China AVS Patent Pool Administration. I lead the negotiations for sub-licensing of the AVS standards. In 2022 I played a leading role in the adoption of the AVS3 video standard by DVB.

28. I have performed expert consulting and expert witness work for patent holders and defendants in patent licensing negotiations and litigation.

F. Curriculum Vitae

29. Additional information concerning my professional publications and presentations in the field of digital video and cases in which I have served as an expert are set forth in my current Curriculum Vitae, a copy of which is attached as Exhibit A. This Curriculum Vitae lists many publications authored or co-authored by me; and lists the cases in which I have testified via depositions and trials.

II. Digital Video Technologies

30. Below I describe the state of digital video coding as of the '303 Patent's earliest claimed priority date (March 31, 2014), and summarize the development of technologies underlying the features and methods recited in claims 1-2.

A. Analog Video

31. Like film, video comprises a rapid sequence of still images that the human vision system integrates into a perception of motion. The “moving images” are often called video frames and video pictures.

32. All-electronic television¹ was invented by Philo Farnsworth in the late 1920s. Farnsworth developed the technique of raster scanning images and produced cameras and CRTs (cathode ray tubes) that could synchronously acquire and reproduce moving images.

33. All-electronic television service was introduced by the BBC in England in November 1936. Broadcasting of television programs continued until interrupted by the Second World War. Television service began in the U.S. in July 1941. These services provided only black-and-white video until 1953, when the NTSC standard in the U.S. was revised to include color. Analog broadcasting continued in the U.S. for almost seventy years. Full-power analog television broadcasting ceased in 2009 under order from the FCC.

B. Digital Video

34. Digitization of analog signals in general was introduced in the late 1940s with the name “pulse code modulation” (PCM). Fundamental sampling theory showed

¹ Electro-mechanical television systems were also introduced.

that a bandwidth-limited analog signal could be sampled at twice the rate of the highest frequency in the band, and still be perfectly reconstructed from such samples. The sample amplitudes could be represented by the value of a number, *e.g.*, a decimal value or a binary value. In the case of image or video data, it has been widespread practice to represent this value by an unsigned 8-bit integer. Analog TV signals that scanned a video frame were digitized into two-dimensional arrays of digital picture elements, known as “pels” or “pixels.” Digital video, like analog video (and film) comprises a rapid sequence of frames.

35. PCM was followed in 1952 by the invention at Bell Labs of differential pulse code modulation, (DPCM), which was immediately applied to video. Because the correlation between pixels in typical scenes was high, it was efficient to code the pixels by successively predicting each pixel from the preceding pixel or pixels. AT&T demonstrated its Picturephone in 1964 at the World’s Fair. Subsequently, commercial Picturephone service was introduced in 1970, using DPCM coding.

36. Digital video was introduced in the professional television studio environment with the CCIR 601 standard that was developed in the early 1980s and published in 1982 (now known as the ITU-R BT.601 standard). The standard covered PCM data formats for the NTSC and PAL video standards and facilitated the development of modern production and post-production studios, involving non-linear editing, digital switchers, and mixers, as well as digital special effects.

37. Digital video in its raw format is highly redundant. Each video frame comprises a regular, rectangular array of pixels. The density of the pixel array is chosen such as to portray the highest spatial resolution enabled by a particular video system. Historically, the earliest systems portrayed 400-500 lines of resolution. The so-called HDTV systems introduced in the 1990s doubled the number of lines to 1000 lines, and the recent so-called 4K systems double the number of lines again to 2000 lines. The imminent next generation of so-called 8K systems will further double the number of lines to 4000 lines, which will approximate the resolution of a 35mm film system.

38. Historically the frame rate was 30 frames per second, but one of the HDTV formats doubled this to 60 frames per second.

39. Transmission or storage of raw digital video data requires huge volumes of data and very fast data rates. Transmitting standard definition color TV with a format of 720x486 pixels and 30 frames/s would require over 250Mbits/s. HDTV would require approximately 5 times as much data and so-called 4K TV (UHDTV) at 60 frames/s would require approximately 40 times as much data.

40. Within a given scene being imaged by a video system, the content is very likely not to contain the highest level of detail except in a few, small areas. While a prescribed density of pixels may accurately describe such high detail areas, that density of pixels is overkill for the rest of the scene to a greater or lesser degree. So the typical digital video frame is spatially redundant. For example, if the video scene contains the

sky, or a uniformly painted wall, adjacent pixels will have almost identical intensity and color.

41. The frame rate of the video is established to smoothly portray the fastest motion the system is designed to support. But such motion may occur only a fraction of the time, and may occur only over a portion of a frame spatially. So the typical video sequence of frames is temporally redundant. For example, sportscasters sitting at a desk demonstrate little motion over only a portion of the scene, but when they cut away to basketball, the action is frenetic and may be scene-wide. In the former case the pixel values will be almost identical from one frame to the next over most of the frame.

C. Video Coding

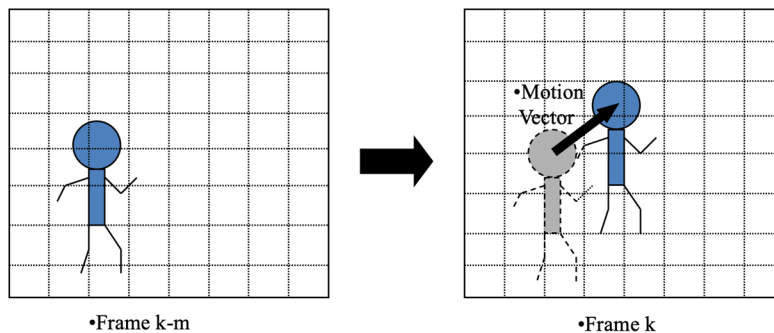
42. Video coding, also known as video compression is the art of removing the spatial and temporal redundancy to enable storage or transmission of digital video data with vastly fewer bits, while preserving the quality of the raw video data to the maximum extent possible. Expressed another way, video coding is the art of representing the information in a video sequence with the least data (i.e., the fewest bits) and the highest reconstructed quality.

43. Video coding is not performed by a single algorithm, rather it comprises a set of video coding tools, each designed to tackle a particular aspect of reducing the redundancy of raw video data. Some video tools tackle the spatial redundancy, while other tools tackle the temporal redundancy. In each case, the tools attempt to identify

and extract the visual information perceived by the human vision system. This information is then encoded into a video bitstream.

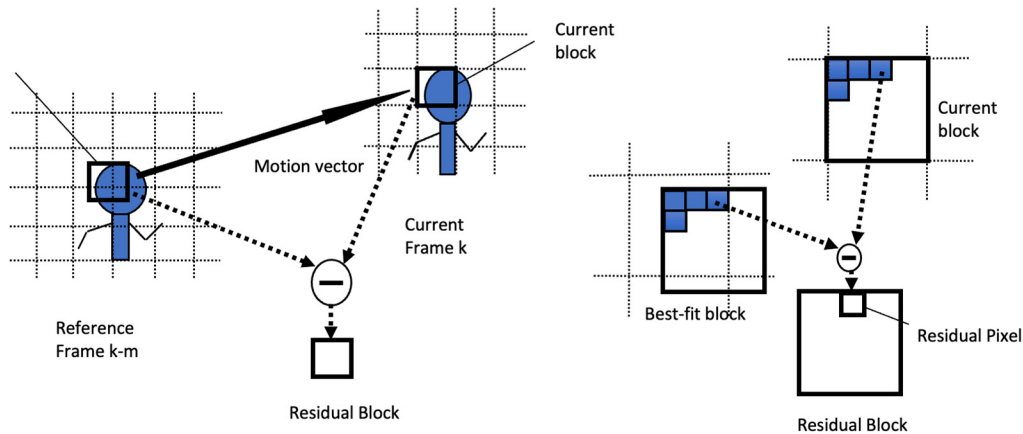
44. The fundamental research for video coding was performed in the 1960s and 1970s. By the end of the 1970s, a so-called hybrid architecture had been developed that remains today as the foundation for almost all video coding standards. Temporal redundancy is removed using block-based motion compensation. This is based on a model of motion in which objects within a scene generally do not change, but rather move from one location in the scene to another. On a block basis, this model works well. The following figures I assembled illustrate the point:

- **Model: Motion consists of translation in the viewing plane.**



45. Thus for each block in a current frame, a search is made to find a best match to the block in a “reference” frame that has already been coded. The best-matched block is used as a prediction for the current block. The difference between the pixel values of the current block and the pixel values of the prediction block is called the residual data. The location of the best match block in the reference frame relative

to the location of the current block is called a motion vector. The residual data block and the motion vector are coded, and typically comprise a small fraction of the amount of data needed to code the current block alone. The following figures I assembled illustrate the point:



46. For blocks that cannot be temporally predicted efficiently, spatial prediction within the same frame is applied. In both cases the blocks of residual data may still contain spatial redundancy within the block. This redundancy is reduced using transform coding techniques.

47. The nature of video data is highly variable, both spatially and temporally. Some video sequences contain high levels of spatial detail. Such detail may span across the whole video frame or may be confined to local areas of the video frame. Some video sequences contain high degrees of movement that also may span across the whole video frame, such as when a camera is panning rapidly across a scene, or may be localized within the video frame. Video coding is performed in a highly adaptive way,

by breaking each video frame into small blocks and adapting the spatial and temporal coding to the characteristics of the data on a local basis.

48. The temporal and spatial information is entropy coded to efficiently represent it and is then formatted into a bitstream.

49. The adaptive coding of video results in a highly variable instantaneous bitrate – it is obvious the volume of data produced when a ball is snapped in a football game is vastly higher than that produced during a time-out. As a result, the entropy coded bitstream at the encoder is buffered in order to smooth out at least some of the variation in data rate before transmission or storage and there is a corresponding buffer at the input to the decoder.

50. Video coding in particular has been characterized by the difficulty and cost involved in implementing coding algorithms for use in mass-market consumer applications. The volume of video data is very large, and the requirement for presenting video data that has been transmitted or retrieved from storage in real-time has been very challenging. As a consequence, while researchers in the field have conceived of sophisticated mathematical algorithms and procedures for video coding for decades, practical implementation has had to wait for semiconductor technology to pass certain thresholds.

D. Video Coding Standards

1. Background

51. Two international bodies are responsible for audiovisual standards – the ITU-T and ISO/IEC. Throughout the last three decades, these bodies have collaborated to develop generations of standards, most frequently publishing identical texts under their own standardization programs.

52. In general, the standardization process has followed the practice of beginning with a requirements gathering stage, leading to a call for proposals and evaluation stage, and then a development stage. In the development stage, the draft standard, and complementary reference software, are iterated from one meeting to the next. Owing to the highly variable nature of video data, the experts propose techniques to develop a particular coding tool and perform experiments using the reference software and selected test sequences to determine whether or not, and under what circumstances, a proposed technique improves the draft coding standard. Frequently, a technique brings improvement under specific, limited circumstances, so it is common practice for a tool to be switched on and off for local regions of the video sequence (spatially and/or temporally), or for parameters of a coding tool to be locally adjusted. In this sense, the development of a standard and attendant adoption of coding tools is an ad hoc process based upon empirical evidence.

53. The structure of standards for audiovisual coding has comprised an audio/speech coding part, video coding part, and a “systems” part. The audio/speech and video parts establish standards normatively specifying precise formats for coded

bitstreams dedicated exclusively to audio/speech and video respectively. These may be called elementary bitstreams. For audiovisual applications such as broadcast television, these elementary bitstreams are combined into a single bitstream by the system part.

2. The MPEG1 Standards

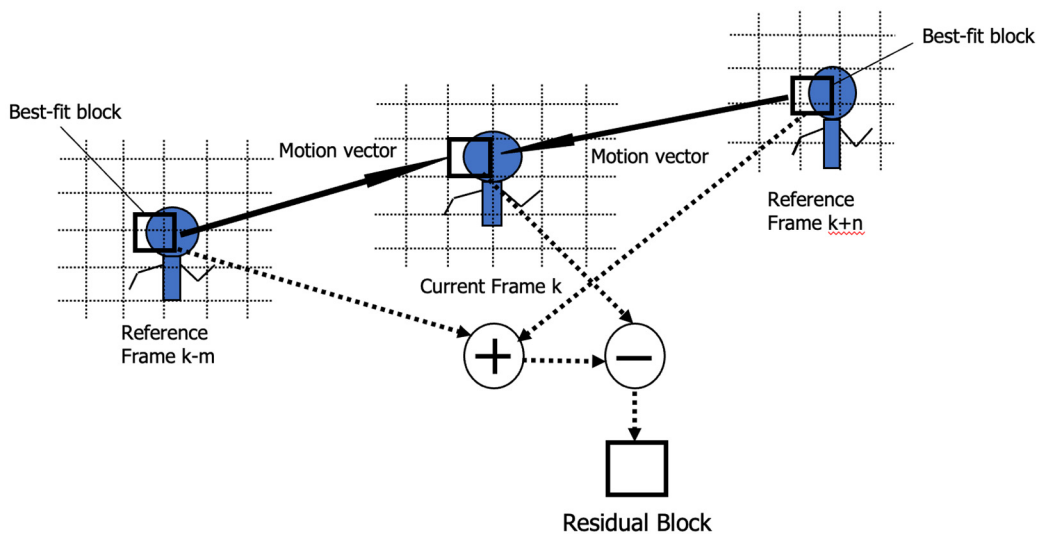
54. The MPEG committee established requirements for the MPEG1 standard and called for proposals in 1989. Many of the resulting proposals were based on the preceding work for the H.261 videoconferencing standard, indeed, MPEG1 may be viewed as an extension of that standard. But for MPEG1, the requirements were focused on a digital replacement for the analog VCR, and included trick play modes such fast-forward, fast rewind and reverse play. Multiple responses to the call proposed groups of pictures as a mechanism to meet these requirements. During the ensuing development of the standard, the Group Of Pictures (GOP) data structure was defined.

2.1.71 group of pictures [video]: A series of one or more coded pictures intended to assist random access. The group of pictures is one of the layers in the coding syntax defined in this part of ISO/IEC 11172.

SAMSUNG-1015, 6.

55. The group of pictures must contain an I-picture, may contain P-pictures and B-pictures and can be any length greater than or equal to one I-picture. SAMSUNG-1015, 15-16. In MPEG1, the coding order and display order of the pictures of the GOP can be different from each other. *Id.*

56. The MPEG1 standard, which is incorporated by reference in this declaration in full, introduced a highly efficient method of temporal coding called bi-directional prediction. This was subsequently adopted by the MPEG2 standard. In this method, multiple video frames are buffered prior to encoding and the coded data for multiple frames are buffered before decoding. Temporal coding is used first to predict and code the blocks in a future frame that typically may be two or three frames ahead in the sequence of buffered frames. This is called a “P-frame.” Then the blocks in an intervening frame are coded by searching for matching blocks in a previous frame and a future frame. This is called a “B-frame.” The corresponding pixel values in the best match blocks are averaged and the averaged block is used to predict the block in the intervening frame. The resultant residual block and the two corresponding motion vectors are coded.



57. The prediction of P-frames by successive forward prediction and B-frames by prediction from adjacent I-frames and P-frames is illustrated in Figure D.2 of the MPEG1 Standard, shown below. It can be seen that intra-coded frame I2 is not predicted from another frame; P-frames P5, P8 and P11 are forward-predicted from I-frames or other P-frames, and pairs of frames I2, P5, P8 and P11 are used to predict the corresponding interstitial B-frames.

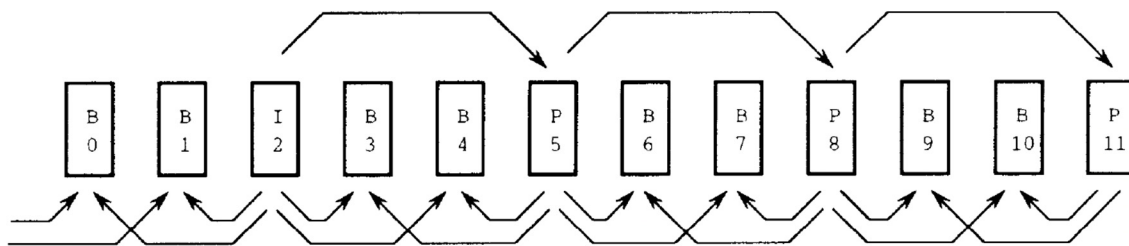


Figure D.2 -- Dependency relationship between I, B, and P-pictures

Figure D.2 illustrates a number of pictures in display order. The arrows show the dependency relationship of the predictive and bidirectionally predictive coded pictures.

SAMSUNG-1015, 54.

58. To the extent Patent Owner interprets a GOP as a “grouping of a series of frames into sequences,” Figure D.2 from the 1993 MPEG1 standard shows it. Not only that, but the GOP from 1993 includes an I-frame (intra-coded picture) and several B- and P-frames (a plurality of inter-prediction coded pictures). The GOP structure claimed by the '303 Patent was described in the MPEG1 standard as early as 1993. An example of Groups of Pictures from MPEG1 in 1993 is reproduced below:

Group of pictures

A **group of pictures** is a series of one or more coded pictures intended to assist random access into the sequence. In the stored bitstream, the first coded picture in a group of pictures is an I-Picture. The order of the pictures in the coded stream is the order in which the decoder processes them in normal playback. In particular, adjacent B-Pictures in the coded stream are in display order. The last coded picture, in display order, of a group of pictures is either an I-Picture or a P-Picture.

The following is an example of groups of pictures taken from the beginning of a video sequence. In this example the first group of pictures contains seven pictures and subsequent groups of pictures contain nine pictures. There are two B-pictures between successive P-pictures and also two B-pictures between successive I- and P-pictures. Picture 'I1' is used to form a prediction for picture '4P'. Pictures '4P' and 'I1' are both used to form predictions for pictures '2B' and '3B'. Therefore the order of pictures in the coded sequence shall be 'I1', '4P', '2B', '3B'. However, the decoder should display them in the order 'I1', '2B', '3B', '4P'.

At the encoder input,

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P

At the encoder output, in the stored bitstream, and at the decoder input,

1	4	2	3	7	5	6	10	8	9	13	11	12	16	14	15	19	17	18	22	20	21	25	23	24
I	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B

where the double vertical bars mark the group of pictures boundaries. Note that in this example, the first group of pictures is two pictures shorter than subsequent groups of pictures, since at the beginning of video coding there are no B-pictures preceding the first I-Picture. However, in general, in display order, there may be B-Pictures preceding the first I-Picture in the group of pictures, even for the first group of pictures to be decoded.

SAMSUNG-1015, 15. “Applications requiring random access, fast-forward playback, or fast and normal reverse playback may use relatively short groups of pictures.

Groups of pictures may also be started at scene cuts or other cases where motion compensation is ineffective. The number of consecutive B-Pictures is variable.” *Id.*

59. As a result of bi-directional prediction, the video frames are encoded and decoded out of order. A direct consequence for the decoder system is a requirement for decoding P-frames before presenting the P-frames. For example in the above illustration, frame P5 is encoded following encoding of frame I2, and then frames B3 and B4 are encoded. This is further illustrated in Figures D.3 and D.4 of the MPEG1 Standard, shown below.

Note that because of the picture dependencies, the bitstream order, i.e. the order in which pictures are transmitted, stored, or retrieved, is not the display order, but rather the order which the decoder requires them to decode the bitstream. An example of a sequence of pictures, in display order, might be:

I	B	B	P	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Figure D.3 -- Typical sequence of pictures in display order

whereas the bitstream order would be as shown below:

I	P	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B
0	3	1	2	6	4	5	9	7	8	12	10	11	15	13	14	18	16	17

Figure D.4 -- Typical sequence of pictures in bitstream order

Because the B-pictures depend on the following (in display order) I or P-picture, the I or P-picture must be transmitted and decoded before the dependent B-pictures.

SAMSUNG-1015, 54.

60. In Figure D.3, the pictures are numbered and sequenced in natural order, corresponding to the order in which the decoded pictures are displayed. In Figure D.4 it can be seen the sequence of picture numbers is not in natural order. The pictures are decoded in this order and then buffered to provide reordering for display.

61. Likewise, to the extent Patent Owner interprets a GOP as a “grouping of a series of frames into sequences,” the GOP structure in Figure D.3 also satisfies the this interpretation of a GOP because D.3 shows a grouping of I-frames (intra-coded pictures) and a plurality of B- and P-frames (inter-prediction coded pictures).

62. It should be noted the MPEG specifications are fully flexible in the number of interstitial B-frames. The number could be 0, 1, 2, ... and may vary dynamically between any pair of predicting frames (i.e., it is not fixed within a GOP).

In the MPEG standards, B-frames are not used to predict other frames. Examples are shown in the following figure:

```

I
I  P  P
I  B  P  B  P
B  B  I  B  P  B  P
B  B  I  B  B  P  B  B  P  B  B  P
B  I  B  B  B  B  P  B  I  B  B  I  I

```

Figure D.17 -- Examples of groups of pictures in display order

SAMSUNG-1015, 68.

63. The number of interstitial B-frames is determined by an encoder and is application-specific. It has been found empirically that two B-frames is optimum for natural video sequences, typified by broadcast television and movies recorded on DVD. It is common practice to have a constant number of B-frames, as shown in the above figures.

64. The MPEG1 standard was technically completed in 1991, and published by the standards setting organization in 1993. *E.g.*, SAMSUNG-1015. By this time, GOP structure was well defined and understood.

3. The MPEG2 Standard

65. The MPEG committee established requirements for the MPEG2 standard and called for proposals in 1991. The requirements were focused first on digital television broadcasting and recording of standard definition video (*e.g.*, NTSC, PAL) at full resolution. As such, MPEG2 comprises an extension of MPEG1 to include coding

of interlaced video. The MPEG2 standard also includes optional coding tools for scalable video coding. *E.g.*, SAMSUNG-1028; SAMSUNG-1016.

66. The MPEG2 standard, officially ISO/IEC 13818-2, which is incorporated by reference in this declaration in full, was technically completed in 1994, and rapidly became massively successful in the global marketplace as the standard for digital television broadcasting and digital video recording – DVD. *E.g.*, SAMSUNG-1028, SAMSUNG-1016.

4. MPEG Scalability

67. In the context of video coding, scalability is a functionality in which a single coded bitstream comprises a hierarchical set of sub-streams representing the video data. The bitstream is hierarchical because a base layer is coded at a given reproducible (decoded, reconstructed) quality level, and also used to predict a layer that is coded at a higher reproducible (decoded, reconstructed) quality level. This process may be iteratively continued to predict further higher quality layers.

68. The complete set represents the video data at the highest quality for reproduction and the highest bitrate for storage and transmission. Subsets of the streams may be stored and transmitted at lower bitrates, leading to lower reproduced quality.

69. A subset may represent the video content at a lower quality, or at a lower spatial resolution, or a subset may represent the video content at a lower temporal resolution, i.e., a lower frame rate.

70. Efficient coding of scalable bitstreams may be provided by coding a “base layer” at a relatively low bitrate with video frames having relatively low quality, spatial resolution, or temporal resolution, then using those frames to predict one or more “enhancement layers” having contemporaneous frames with higher quality, higher spatial resolution, or additional frames to provide higher temporal resolution. The concept is the enhancement layer frames are highly correlated with the base layer frames, so coding only the difference between them requires few bits.

71. In the MPEG standards, B-frames inherently support temporal scalability because they are not used to predict other frames. Examples of practical applications for such temporal scalability include a heterogeneous network in which a node on a broadband network interfaces to a narrow band network, at which a full resolution, full bandwidth video bitstream may be forwarded by stripping out the bits used to code the B-frames. Another example is a mobile receiver having limited computational power, that decodes and presents only the I-frames and P-frames.

72. The MPEG2 standard supports three types of scalability – SNR scalability (i.e. quality scalability), spatial scalability and temporal scalability. SAMSUNG-1016, 146. Section D.7.1.2 describes these video coding tools. The SNR scalability is

summarized as : “The aim of SNR scalability is primarily to provide a mechanism for transmission of a two layer service, these two layers providing the same picture resolution but different quality level.” In brief, section F.7.1.2.1 explains how SNR scalability functions in the following way (emphasis added):

In the lower layer, the encoding is similar to the non scalable situation in terms of decisions, adaptive quantization, buffer regulation. The intra or error prediction macroblocks are DCT transformed. The coefficients are then quantized using a first rather coarse quantizer. The quantized coefficients are then VLC coded and sent together with the required side information (macroblock_type, motion vectors, coded_block_pattern()).

In parallel, the quantized DCT coefficients coming from the lower layer, are de-quantized. The residual error between the coefficients and the de-quantized coefficients is then re-quantized, using a second finer quantizer. The resulting refinement coefficients are VLC coded and form the additional enhancement layer, together with a marginal amount of side information (quantiser_scale_code, coded_block_pattern() ...).

SAMSUNG-1016, 146.

73. The spatial scalability in MPEG2 is implemented by a variant of the SNR scalability in which the video may be decoded as a 2:1 reduced resolution.

74. Temporal scalability in MPEG2 is illustrated in the figure below and summarized in section D.7.4 as follows:

A two layer temporally scalable coding structure consisting of a base and an enhancement layer is shown in Figure F.1. Consider video input at full temporal rate to temporal demultiplexer; in our example it is temporally demultiplexed to form two video sequences, one input to the base layer encoder and the other input to the enhancement layer encoder. The base layer encoder is a non-

hierarchical encoder operating at half temporal rate, the enhancement layer encoder is like a MAIN profile encoder and also operates at half temporal rate except that it uses base layer decoded pictures for motion compensated prediction. The encoded bitstreams of base and enhancement layers are multiplexed as a single stream in the systems multiplexer. The systems demultiplexer extracts two bitstreams and inputs corresponding bitstreams to base and enhancement layer decoders. The output of the base layer decoder can be shown standalone at half temporal rate or after multiplexing with enhancement layer decoded frames and shown at full temporal rate.

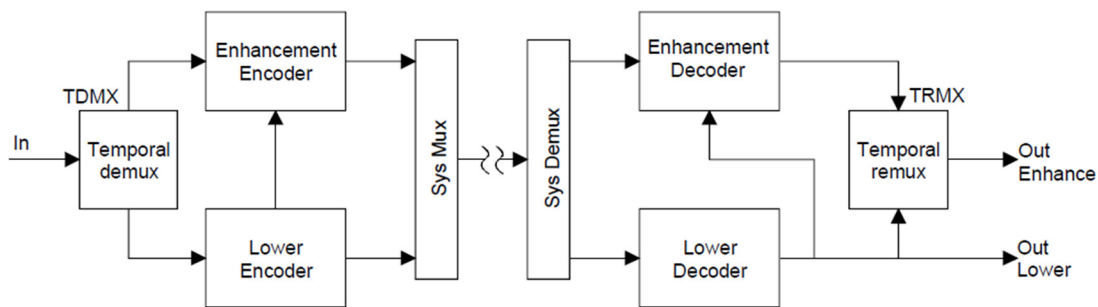


Figure D.1 – A two-layer codec structure for temporal scalability

SAMSUNG-1016, 147.

75. The MPEG4 Part 2 Visual standard includes fine grain scalability “FGS.”
E.g., SAMSUNG-1009 at , A comprehensive overview can be found in Radha et al, SAMSUNG-1024, which is incorporated by reference in this declaration in full. In brief, this comprises the following:

In order to meet the requirements outlined in the previous section, FGS encoding is designed to cover any desired bandwidth range while maintaining a very simple scalability structure. ... the FGS structure consists of only two layers: a base-layer coded at a bitrate and a single enhancement-layer coded using a fine-granular (or embedded) scheme to a maximum bitrate of R_e . This structure provides a very efficient, yet

simple, level of abstraction between the encoding and streaming processes. The encoder only needs to know the range of bandwidth $[R_{\min} - R_b, R_{\max} = R_e]$ over which it has to code the content, and it does not need to be aware of the particular bitrate the content will be streamed at. The streaming server on the other hand has a total flexibility in sending any desired portion of any enhancement layer frame (in parallel with the corresponding base layer picture), without the need for performing complicated real-time rate control algorithms. This enables the server to handle a very large number of unicast streaming sessions and to adapt to their bandwidth variations in real-time. On the receiver side, the FGS framework adds a small amount of complexity and memory requirements to any standard motion- compensation based video decoder.

...

... the FGS framework requires two encoders, one for the base-layer and the other for the enhancement layer. The base-layer can be compressed using any motion-compensation video encoding method. Naturally, the DCT-based MPEG-4 video standard is a good candidate for the base-layer encoder due to its coding efficiency especially at low bitrates.

Using the DCT transform at both the base and enhancement layers enables the encoder to perform a simple residual computation of the FGS enhancement-layer. Each DCT FGS-residual frame consists of N_{BP} bitplanes:

$$N_{BP} = \lfloor \log_2 (|C|_{\max}) \rfloor + 1$$

Where $|C|_{\max}$ is the maximum DCT (magnitude) value of the residual frame under consideration.³ After identifying $|C|_{\max}$ and the corresponding N_{BP} , the FGS enhancement-layer encoder scans the residual signal using the traditional zig-zag scanning method starting from the most significant bitplane BP(1) and ending at the least significant bitplane BP(N_{BP}). Every bitplane consists of

nonoverlapping macroblocks (MB's), and each MB includes four luminance () blocks and two chroma blocks (and). Run-length codes are used for (lossless) entropy-coding of the zeros and ones in each bitplane block. This process generates variable length codes that constitute the FGS compressed bitstream. A special "all-zero blocks" code is used when all six bitplane-blocks (within a given bitplane-macroblock) do not have any bits with a value of one. At the receiver side, the FGS bitstream is first decoded by a variable length decoder (VLD). Due to the embedded nature of the FGS stream, the VLD re-generates the DCT residual bitplanes starting from the most significant bitplane toward the least significant one. Moreover, due to the type of scanning used by the FGS encoder, it is possible that the decoder does not receive all of the bitplane-blocks that belong to a particular bitplane. Any bitplane block not received by the decoder can be filled with zero values. The resulting DCT residual is then inverse-transformed to generate the SNR residual pixels. These residual pixels are then added to the base-layer decoder output to generate the final enhanced scalable video.

Radha et al., SAMSUNG-1024.

E. The H.263 Standard

76. H.263 is a standard developed by ITU-T for videoconferencing applications. The work commenced in 1993, leading to three versions of the standard in 1996, 1998 and 2001. The first version is an evolution of the earlier H.261 standard. The second version, which is incorporated by reference in this declaration in full, included a number of optional annexes that are relevant to the current matter. See, *e.g.*, SAMSUNG-1020.

77. For example, SAMSUNG-1020, Annex O Specifies three scalability modes, similar to MPEG2: SNR, spatial and temporal. In brief, section O.1.1 discloses temporal scalability as follows:

Temporal scalability is achieved using bidirectionally predicted pictures, or B-pictures. B-pictures allow prediction from either or both a previous and subsequent reconstructed picture in the reference layer. This property generally results in improved compression efficiency as compared to that of P-pictures. B-pictures (and the B-part of PB- or Improved PB-frames) are not used as reference pictures for the prediction of any other pictures. This property allows for B-pictures to be discarded if necessary without adversely affecting any subsequent pictures, thus providing temporal scalability. Figure O.1 illustrates the predictive structure of P- and B-pictures.

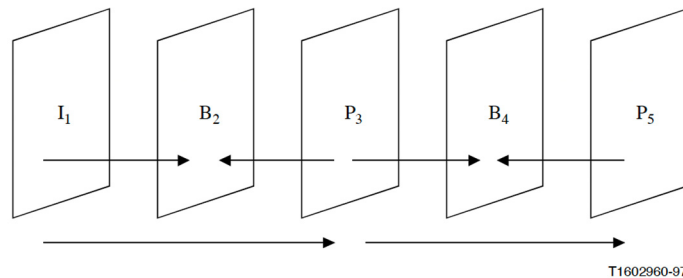


Figure O.1/H.263 – Illustration of B-picture prediction dependencies

SAMSUNG-1020, 102-103.

78. SNR scalability is disclosed in section O.1.2 as follows:

The other basic method to achieve scalability is through spatial/SNR enhancement. Spatial scalability and SNR scalability are equivalent except for the use of interpolation as is described shortly. Because compression introduces artifacts and distortions, the difference between a reconstructed picture and its original in the encoder is (nearly always) a nonzero-valued picture, containing what can be called the coding error. Normally, this coding error is lost at the encoder and never recovered. With SNR scalability, these

coding error pictures can also be encoded and sent to the decoder, producing an enhancement to the decoded picture. The extra data serves to increase the signal-to-noise ratio of the video picture, and hence, the term SNR scalability. Figure O.2 illustrates the data flow for SNR scalability. The vertical arrows from the lower layer illustrate that the picture in the enhancement layer is predicted from a reconstructed approximation of that picture in the reference (lower) layer. If prediction is only formed from the lower layer, then the enhancement layer picture is referred to as an EI-picture. It is possible, however, to create a modified bidirectionally predicted picture using both a prior enhancement layer picture and a temporally simultaneous lower layer reference picture. This type of picture is referred to as an EP-picture or "Enhancement" P-picture. The prediction flow for EI- and EP-pictures is shown in Figure O.2. (Although not specifically shown in Figure O.2, an EI-picture in an enhancement layer may have a P-picture as its lower layer reference picture, and an EP-picture may have an I-picture as its lower-layer enhancement picture.) For both EI- and EP-pictures, the prediction from the reference layer uses no motion vectors. However, as with normal P-pictures, EP-pictures use motion vectors when predicting from their temporally prior reference picture in the same layer.

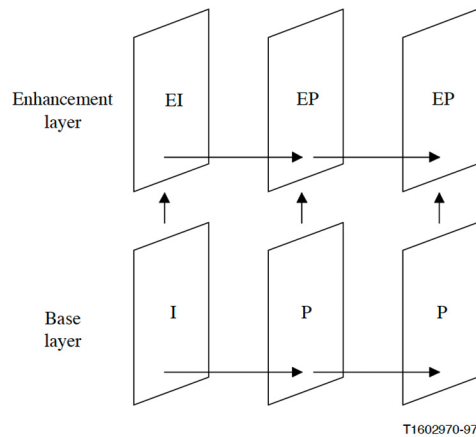


Figure O.2/H.263 – Illustration of SNR scalability

SAMSUNG-1020, 103-104.

79. Spatial scalability is disclosed in section O.1.3 as follows:

The third and final scalability method in the Temporal, SNR, and Spatial Scalability mode is spatial scalability, which is closely related to SNR scalability. The only difference is that before the picture in the reference layer is used to predict the picture in the spatial enhancement layer, it is interpolated by a factor of two either horizontally or vertically (1-D spatial scalability), or both horizontally and vertically (2-D spatial scalability).

Other than requiring an upsampling process to increase the size of the reference layer picture prior to its use as a reference for the encoding process, the processing and syntax for a spatial scalability picture is functionally identical to that for an SNR scalability picture. Since there is very little syntactical distinction between pictures using SNR scalability and pictures using spatial scalability, the pictures used for either purpose are called EI- and EP-pictures. The picture in the base layer which is used for upward prediction in an EI- or EP-picture may be an I-picture, a P-picture, or the P-part of a PB- or Improved PB-frame (but shall not be a B-picture or the B-part of a PB- or Improved PB-frame).

SAMSUNG-1020, 104-105.

80. Spatial scalability is also provided by two other Annexes. Annex P specifies a Reference Picture Resampling, while Annex Q specifies a Reduced Resolution Update. Annex P Section P.1 Introduction of SAMSUNG-1020 explains:

This annex describes the use and syntax of a resampling process which can be applied to the previous decoded reference picture in order to generate a "warped" picture for use in predicting the current picture. This resampling syntax can specify the relationship of the current picture to a prior picture having a different source format, and can also specify a "global motion" warping alteration of the shape, size, and location of the prior picture with respect to the current picture. In particular, the Reference Picture Resampling mode can be used to adaptively alter the resolution of pictures during encoding.

SAMSUNG-1020, 116.

81. Section P.5 Factor-of-4 Resampling further specifies:

Factor-of-4 resampling, which converts both the horizontal and vertical size of the picture by a factor of 2 or $\frac{1}{2}$, is a special case of the resampling algorithm described in P.3. The simplified description of the resampling algorithm for this special case is provided in this subclause.

SAMSUNG-1020, 126.

82. Annex Q Section Q.1 Introduction explains:

This annex describes an optional Reduced-Resolution Update mode of this Recommendation. The capability of this mode is signaled by external means (for example, Recommendation H.245). The use of this mode is indicated in the PLUSPTYPE field of the picture header. The Reduced-Resolution Update mode is expected to be used when encoding a highly active scene, and provides the opportunity to increase the coding picture rate while maintaining sufficient subjective quality. This mode allows the encoder to send update information for a picture that is encoded at a reduced resolution, while preserving the detail in a higher resolution reference image to create a final image at the higher resolution.

SAMSUNG-1020, 129.

83. All of these Annexes can be used in conjunction with Annex N, which specifies the use of multiple reference pictures and the capability to recover from a transmission error by transmitting a new reference picture (or portion of a new reference picture) when the decoder receives a corrupted picture. Section N.2 provides an introduction:

The source coder of this mode is shown in generalized form in Figure N.1. This figure shows a structure which uses a number of picture memories. The source coder may select one of the picture memories to suppress the temporal error propagation due to the

inter-frame coding. The information to signal which picture is selected for prediction is included in the encoded bitstream.

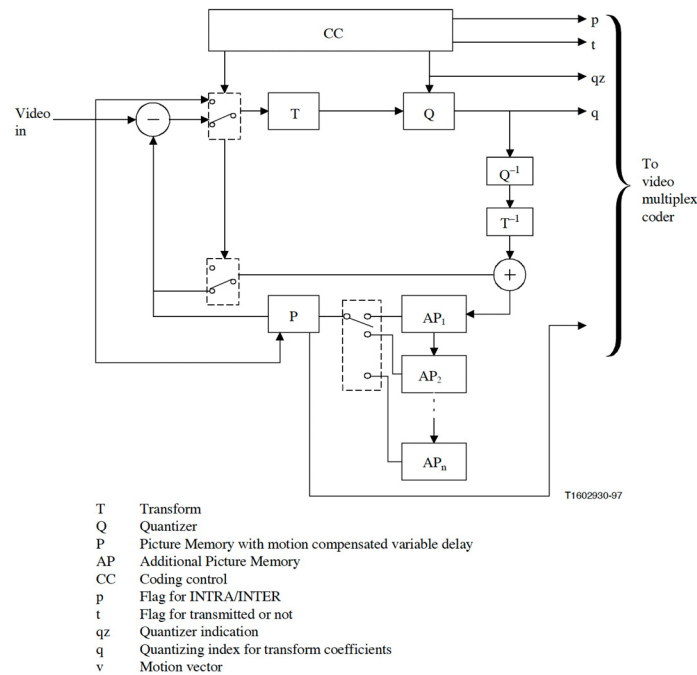


Figure N.1/H.263 – Source coder for NEWPRED

SAMSUNG-1020, 97.

84. Upon detecting an error in a received picture, the decoder may use a backchannel to signal the encoder to send a new picture. The encoder may choose to send an I-picture, so the decoder can reinitialize its temporal prediction of pictures and recover from the error propagation. Because I-pictures require a significant number of transmitted bits, the encoder may combine Annex N with any of the mechanisms in Annexes O, P, and Q to transmit the I-picture at 2:1 reduced resolution. (Transmitting a full-resolution I-picture may require 5-10 times as many bits as a P-picture meaning it could take 5-10 picture display periods. The encoder and decoder rate buffers are sized to manage the variable rate at which pictures are transmitted.)

85. The new reference picture is signaled using its temporal reference. In the normal mode of operation, the most recent temporally previous picture is used as the reference picture. Section N.4.1 specifies a flag Temporal Reference for Prediction Indicator (TRPI), which when set causes a Temporal Reference for Prediction (TRP) to be transmitted. Section N.4.1.4 specifies:

When present (as indicated in TRPI), TRP indicates the Temporal Reference which is used for prediction of the encoding, except in the case of B-pictures and the B-picture part of an Improved PB-frame. For B-pictures or the B-picture part of an Improved PB-frame, the picture having the temporal reference TRP is used for the prediction in the forward direction. (Prediction in the reverse temporal direction always uses the immediately temporally subsequent picture.) TRP is a ten-bit number. If a custom picture clock frequency was not in use for the reference picture, the two MSBs of TRP are zero and the LSBs contain the eight-bit TR found in the picture header of the reference picture. If a custom picture clock frequency was in use for the reference picture, TRP is a ten-bit number consisting of the concatenation of ETR and TR from the reference picture header. When TRP is not present, the most recent temporally previous anchor picture shall be used for prediction, as when not in the Reference Picture Selection mode. TRP is valid until the next PSC, GSC, or SSC.

SAMSUNG-1020, 99.

86. In a videoconferencing application the round-trip delay should be less than 400ms to provide for smooth conversation. At a picture rate of 30 pictures/s the one-way delay should be less than 6 picture periods (200ms). So this represents the time available for the encoder to respond to the backchannel signal from the decoder.

It is clear that encoding and sending a reduced resolution I-picture is the safest method to avoid glitches in the presentation. Section N.5 specifies the decoder process:

The decoder of this mode may need an additional number of picture memories to store the correctly decoded video signals and their Temporal Reference (TR) information. The decoder uses the stored picture for which the TR is TRP as the reference picture for inter-frame decoding instead of the last decoded picture, if the TRP field exists in the forward-channel data. When the picture for which the TR is TRP is not available at the decoder, the decoder may send a forced INTRA update signal to the encoder by external means (for example, Recommendation H.245). Unless a different frame storage policy is negotiated by external means, correctly decoded video picture segments shall be stored into memory for use as later reference pictures on a first-in, first-out basis as shown in Figure N.1 (except for B-pictures, which are not used as reference pictures), and video picture segments which are detected as having been incorrectly decoded should not replace correctly decoded ones in this memory area.

An Acknowledgment message (ACK) and a Non-Acknowledgment message (NACK) are defined as back-channel messages. An ACK may be returned when the decoder decodes a video picture segment successfully. NACKs may be returned when the decoder fails to decode a video picture segment, and may continue to be returned until the decoder gets the expected forward channel data which includes the requested TRP or an INTRA update. Which types of message shall be sent is indicated in the RPSMF field of the picture header of the forward-channel data.

SAMSUNG-1020, 101-102.

F. The H.264 Standard

87. H.264 is a standard jointly developed by ITU-T for videoconferencing applications and ISO for consumer video applications. (The ISO designation is 14496-10, i.e., MPEG4 Part 10 also known as AVC – Advanced Video Codec, SAMSUNG-

1027.) The work commenced in 1998 in the ITU-T workgroup VCEG (Video Coding Experts Group), with MPEG joining in December 2001 to form the Joint Video Team – JVT. The first version was published in May 2003. Scalable Video Coding (SVC) was included from version 3 in November 2007 (ITU-T Rec. H.264 (11/2007), SAMSUNG-1009, which is incorporated by reference in this declaration in full.

88. H.264/AVC provides a significant improvement in video coding performance over MPEG2 and is still the most widely used video coding standard worldwide. Notably it is used for High Definition Television broadcasting (HDTV).

89. For example, Schwarz IEEE paper “Overview of the scalable video coding extension of the H. 264 AVC standard” by three of the principle contributors to the H.264 scalability specification, SAMSUNG-1018, which is incorporated by reference in this declaration in full, provides a thorough overview of the practical applications facilitated by the H.264 improvements.

90. Like the earlier standards, H.264/AVC supports the three types of scalability – temporal, spatial and quality (SNR).

91. Temporal scalability may be implemented using hierarchical B-Pictures. In H.264/AVC B-pictures can be used as reference pictures. Figure 1 of the paper illustrates different modes:

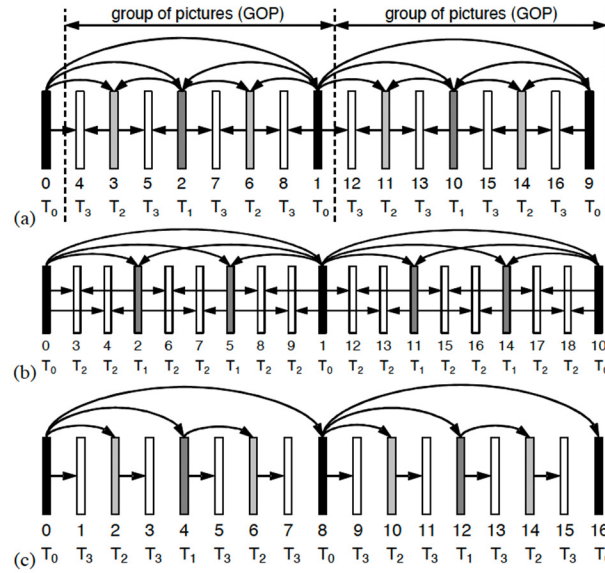


Fig. 1. Hierarchical prediction structures for enabling temporal scalability: (a) coding with hierarchical B pictures, (b) non-dyadic hierarchical prediction structure, (c) hierarchical prediction structure with a structural encoder/decoder delay of zero. The numbers directly below the pictures specify the coding order, the symbols T_k specify the temporal layers with k representing the corresponding temporal layer identifier.

SAMSUNG-1018, FIG. 1.

92. Note the pictures are arranged as GOPs in which base-layer pictures demarcate the GOP boundaries.

Temporal scalability with dyadic temporal enhancement layers can be very efficiently provided with the concept of hierarchical B pictures [28][29] as illustrated in Fig. 1a2. The enhancement layer pictures are typically coded as B pictures, where the reference picture lists 0 and 1 are restricted to the temporally preceding and succeeding picture, respectively, with a temporal layer identifier less than the temporal layer identifier of the predicted picture. Each set of temporal layers $\{T_0, \dots, T_k\}$ can be decoded independently of all layers with a temporal layer identifier $T > k$. In the following, the set of pictures between two successive pictures of the temporal base layer together with the succeeding base layer picture is referred to as a group of pictures (GOP). Although the described prediction structure with hierarchical B pictures provides temporal scalability and also shows excellent coding efficiency as will be demonstrated later, it represents a special case. In general, hierarchical prediction

structures for enabling temporal scalability can always be combined with the multiple reference picture concept of H.264/AVC. This means that the reference picture lists can be constructed by using more than one reference picture, and they can also include pictures with the same temporal level as the picture to be predicted. Furthermore, hierarchical prediction structures are not restricted to the dyadic case. As an example, Fig. 1b illustrates a non-dyadic hierarchical prediction structure, which provides 2 independently decodable sub-sequences with 1/9-th and 1/3-rd of the full frame rate. It should further be noted that it is possible to arbitrarily modify the prediction structure of the temporal base layer, *e.g.*, in order to increase the coding efficiency. The chosen temporal prediction structure does not need to be constant over time. Note that it is possible to arbitrarily adjust the structural delay between encoding and decoding a picture by restricting motion-compensated prediction from pictures that follow the picture to be predicted in display order. As an example, Fig. 1c shows a hierarchical prediction structure, which does not employ motion-compensated prediction from pictures in the future. Although this structure provides the same degree of temporal scalability as the prediction structure of Fig. 1a, its structural delay is equal to zero compared to 7 pictures for the prediction structure in Fig. 1a. However, such low-delay structures typically decrease coding efficiency.

SAMSUNG-1018.

93. Spatial scalability in H.264/AVC is an enhanced version of the spatial scalability in the earlier standards. It is illustrated in Figure 4, which includes a combination with temporal scalability:

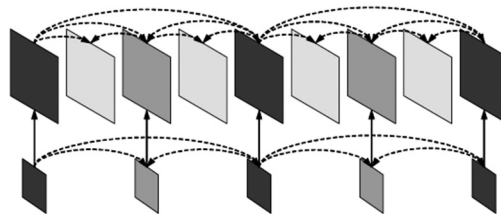


Fig. 4. Multi-layer structure with additional inter-layer prediction for enabling spatial scalable coding.

For supporting spatial scalable coding, SVC follows the conventional approach of multi-layer coding, which is also used in H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual. Each layer corresponds to a supported spatial resolution and is referred to by a spatial layer or dependency identifier D. The dependency identifier D for the base layer is equal to 0, and it is increased by 1 from one spatial layer to the next. In each spatial layer, motion-compensated prediction and intra prediction are employed as for single-layer coding. But in order to improve coding efficiency in comparison to simulcasting different spatial resolutions, additional so-called inter-layer prediction mechanisms are incorporated as illustrated in Fig. 4.

In order to restrict the memory requirements and decoder complexity, SVC specifies that the same coding order is used for all supported spatial layers. The representations with different spatial resolutions for a given time instant form an access unit and have to be transmitted successively in increasing order of their corresponding spatial layer identifiers D. But as illustrated in Fig. 4, lower layer pictures do not need to be present in all access units, which makes it possible to combine temporal and spatial scalability.

SAMSUNG-1018.

94. Quality scalability in the H.264/AVC standard is provided in several modes, one of which is called coarse grain scalability – CGS. The paper explains:

Quality scalability can be considered as a special case of spatial scalability with identical picture sizes for base and enhancement layer. As already mentioned in sec. V.B, this case is supported by the general concept for spatial scalable coding and it is also referred to as coarse-grain quality scalable coding (CGS). The same inter-layer prediction mechanisms as for spatial scalable coding are employed, but without using the corresponding upsampling operations and the inter-layer deblocking for intra-coded reference layer macroblocks. Furthermore, the inter-layer intra and residual prediction are directly performed in the transform domain. When utilizing inter- layer prediction for coarse-grain quality scalability in

SVC, a refinement of texture information is typically achieved by requantizing the residual texture signal in the enhancement layer with a smaller quantization step size relative to that used for the preceding CGS layer.

SAMSUNG-1018.

95. The quality scalable modes deployed by the legacy standards and H.264/AVC are illustrated in Figure 8:

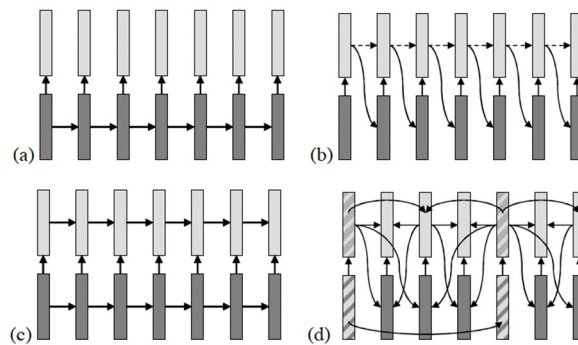


Fig. 8. Various concepts for trading off enhancement layer coding efficiency and drift for packet-based quality scalable coding: (a) base layer only control, (b) enhancement layer only control, (c) two-loop control, (d) key picture concept of SVC for hierarchical prediction structures, where key pictures are marked by the hatched boxes.

SAMSUNG-1018.

96. Schwarz explains the differences in detail, but in brief, the issues include drift in the feedback loop of the decoder, under transmission error conditions, and the cost of providing error resilience in terms of coding efficiency.

97. H.264 includes an optional tool for switching efficiently between contemporaneous coded streams. For example, two streams may comprise the same content with sequences of the pictures coded at different bitrates, and the functionality is desired to switch from one bitrate to the other bitrate seamlessly. While that could

be accomplished with periodic I-pictures, such pictures would require a significant number of coded bits and attendant loss of coding efficiency. H.264 provides a tool at the Slice level called SP Slices, in which the slices are predictively coded with fewer bits in such a way that either stream may continue error-free decoding from the SP Slice. Of course, such a Slice may cover the entire Picture if desired. This tool is disclosed in a white paper - H.264 / MPEG-4 Part 10 : SP and SI slices (SAMSUNG-1025). In brief:

2. SP and SI slices

SP and SI slices are specially-coded slices that enable (among other things) efficient switching between video streams and efficient random access for video decoders. A common requirement is for a video decoder to switch between one of several encoded streams. For example, the same video material is coded at multiple bitrates for transmission across the Internet; a decoder attempts to decode the highest-bitrate stream it can receive but may require to switch automatically to a lower-bitrate stream if the data throughput drops.

Example: A decoder is decoding Stream A and wants to switch to decoding Stream B (Figure 2-1). For simplicity, assume that each frame is encoded as a single slice and predicted from one reference (the previous decoded frame). After decoding P-slices A0 and A1, the decoder wants to switch to Stream B and decode B2, B3 and so on. If all the slices in Stream B are coded as P-slices, then the decoder will not have the correct decoded reference frame(s) required to reconstruct B2 (since B2 is predicted from the decoded frame B1 which does not exist in stream A). One solution is to code frame B2 as an I-slice. Because it is coded without prediction from any other frame, it can be decoded independently of preceding frames in stream B. The decoder can therefore switch between stream A and stream B as shown in Figure 2-1. Switching can be accomplished by inserting an I-slice at regular intervals in the coded sequence to create “switching points”. However, an I-slice is likely to contain much more coded data than a P-slice and the result is a peak in the coded bitrate at each switching point.

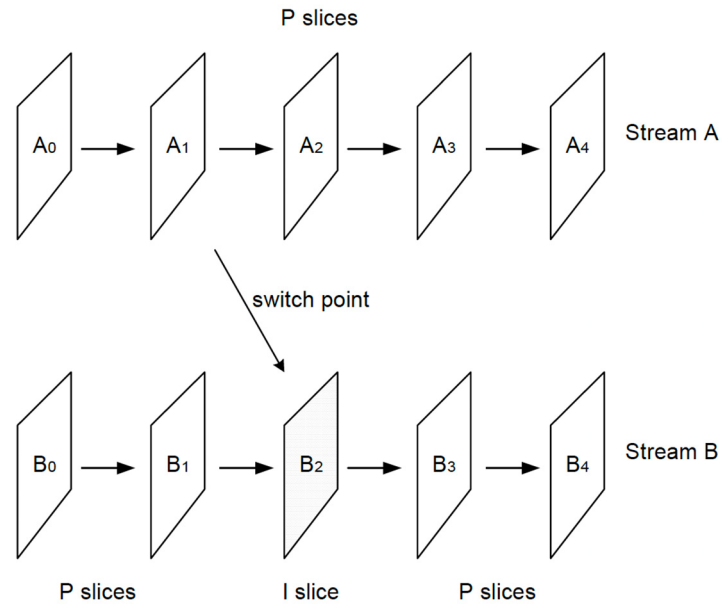


Figure 2-1 Switching streams using I-slices

SP-slices are designed to support switching without the increased bitrate penalty of I-slices. Figure 2-2 shows how they may be used. At the switching point (frame 2 in each sequence), there are now 3 coded SP-slices. These are coded using motion compensated prediction (making them more efficient than I-slices). SP-slice A₂ can be decoded using reference frame A₁; SP-slice B₂ can be decoded using reference frame B₁. The key to the switching process is SP-slice AB₂ (known as a switching SP-slice): this is created in such a way that it can be decoded using motion-compensated reference frame A₁, to produce decoded frame B₂ (i.e. the decoder output frame B₂ is identical whether decoding B₁ followed by B₂ or A₁ followed by AB₂). An extra SP-slice is required at each switching point (and in fact another SP-slice, BA₂, would be required to switch in the other direction) but this is still more efficient than encoding frames A₂ and B₂ as I-slices.

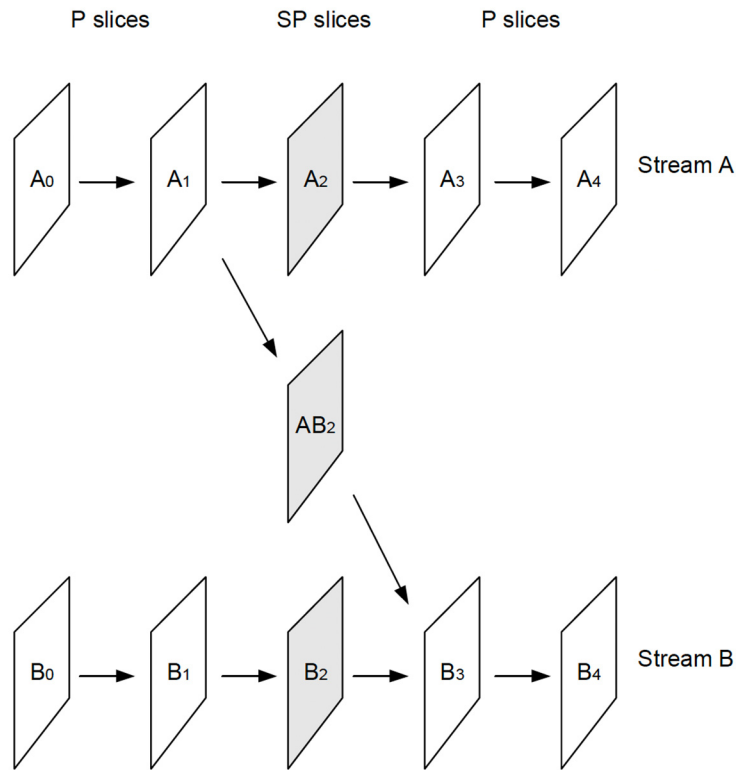


Figure 2-2 Switching streams using SP-slices

SAMSUNG-1025.

98. Of course one way to encode Stream A at one bitrate and Stream B at another bitrate is to use a different resolution for each stream.

99. A comprehensive overview can be found in Karczewicz M, et al. SAMSUNG-1026, which is incorporated by reference in this declaration in full.

100. H.264 specifies a mechanism called the Network Abstraction Layer – NAL – for encapsulating the coded video data and associated parametric data into data units that may be stored or transmitted.

3.89 NAL unit: A syntax structure containing an indication of the type of data to follow and *bytes* containing that data in the form of an RBSP interspersed as necessary with *emulation prevention bytes*.

SAMSUNG-1009 clause 3.97

Scalable video coding is specified in Annex G allowing the construction of bitstreams that contain sub-bitstreams that conform to this Specification. For temporal bitstream scalability, i.e., the presence of a sub-bitstream with a smaller temporal sampling rate than the bitstream, complete access units are removed from the bitstream when deriving the sub-bitstream. In this case, high-level syntax and inter prediction reference pictures in the bitstream are constructed accordingly. For spatial and quality bitstream scalability, i.e., the presence of a sub-bitstream with lower spatial resolution or quality than the bitstream, NAL units are removed from the bitstream when deriving the sub-bitstream. In this case, inter-layer prediction, i.e., the prediction of the higher spatial resolution or quality signal by data of the lower spatial resolution or quality signal, is typically used for efficient coding.

SAMSUNG-1009.

III. Relevant Legal Standards

101. I have been asked to provide my opinions as to whether claims 1-4 of the '303 Patent would have been obvious to a person of ordinary skill in the art as of the earliest claimed priority date of the '303 patent (March 31, 2014) ("Critical Date").

102. I am an engineer by training and profession. The opinions I express in this declaration involve the application of my technical knowledge and experience to the evaluation of certain prior art with respect to the '303 Patent. In addition, I understand that the following legal principles apply.

103. It is my understanding that, in determining whether claims of the '303 patent are obvious in this proceeding, the claim terms are generally given their ordinary and customary meaning as understood by a person of ordinary skill in the relevant art.

A person of ordinary skill in the art would read the claim terms in the context of the entire patent specification in which they appear, as well as the prosecution history of the patent.

104. It is my understanding that a claim is unpatentable under 35 U.S.C. § 103 if the claimed subject matter as a whole would have been obvious to a person of ordinary skill in the art at the time of the alleged invention. I also understand that an obviousness analysis takes into account the scope and content of the prior art, the differences between the claimed subject matter and the prior art, and the level of ordinary skill in the art at the time of the invention.

105. In determining the scope and content of the prior art, it is my understanding that a reference is considered relevant prior art if it falls within the field of the inventor's endeavor. In addition, a reference is prior art if it is reasonably pertinent to the particular problem with which the inventor was involved. A reference is reasonably pertinent if it logically would have commended itself to an inventor's attention in considering his problem. If a reference relates to the same problem as the claimed invention, that supports use of the reference as prior art in an obviousness analysis.

106. To assess the differences between prior art and the claimed subject matter, it is my understanding that 35 U.S.C. § 103 requires the claimed invention to be considered as a whole. This "as a whole" assessment involves showing that one of

ordinary skill in the art at the time of invention, confronted by the same problems as the inventor and with no knowledge of the claimed invention, would have selected the elements from the prior art and combined them in the claimed manner.

107. It is my further understanding that several rationales may be applied for combining references or modifying a reference to show obviousness of claimed subject matter. These rationales include: combining prior art elements according to known methods to yield predictable results; simple substitution of one known element for another to obtain predictable results; a predictable use of prior art elements according to their established functions; applying a known technique to a known device (method or product) ready for improvement to yield predictable results; choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success; and some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify a prior art reference or to combine prior art teachings to arrive at the claimed invention.

IV. Person of Ordinary Skill in the Art

108. It is my understanding that when interpreting the claims of the '303 patent I must do so based on the perspective of one of ordinary skill in the art at the time of the alleged invention ("POSITA"). For purposes of my analysis in this declaration, I have taken the earliest claimed priority date of the '303 patent (March 31, 2014) as the date of the alleged invention ("Critical Date"). I understand that the factors considered

in determining the ordinary level of skill in a field of art include the level of education and experience of persons working in the field; the types of problems encountered in the field; the teachings of the prior art, and the sophistication of the technology at the time of the alleged invention. I understand that a POSITA is not a specific real individual, but rather is a hypothetical individual having the qualities reflected by the factors above. I understand that a POSITA would also have knowledge from the teachings of the prior art, including the art cited below.

109. Taking these factors into consideration, it is my opinion that one of ordinary skill in the art in the field of digital video coding as of the Critical Date, would have had a bachelor's degree in computer science, electrical engineering, computer engineering, or a related field, with 2-3 years of experience in digital video processing, or a related field. Additional graduate education could substitute for professional experience, or additional professional experience could substitute for formal education.

110. I have possessed the qualifications of a POSITA since the Critical Date of the '303 patent, and long before.

V. Overview of the Subject Patent US9986303

111. The '303 patent is titled Video Image Coding Data Transmitter, Video Image Coding Data Transmission Method, Video Image Coding Data Receiver, and Video Image Coding Data Transmission and Reception System. It was filed March 13,

2015, and derived from a Japanese application, filed March 31, 2014. '303 Patent, Cover.

A. Overview

1. Field of Art

112. The stated field of art is “video image coding data transmission technology and, more particularly, to a video image coding data transmission technology for transmitting video image coding data in real time.” '303 Patent, 1:17-21. The stated purpose of the 303 patent is “to provide a video image transmission technology capable of achieving real-time video image transmission and viewing of high-quality video images at the same time.” '303 Patent, 1:64-67.

2. Background - Problem statement, Admitted Prior Art

113. The patent discloses the extensive work on video coding standardization since the 1990s. Two sets of standards are cited:

- MPEG-2 Video (ISO/IEC 13818-2)
- MPEG-4 AVC/H.264 (14496-10 in ISO IEC, H.264 in ITU-T)

3. Summary of the disclosed subject matter of the '303 Patent

114. In brief, the purpose of the '303 Patent's disclosed subject matter is summarized as:

“to provide a video image transmission technology capable of achieving real-time video image transmission and viewing of high-quality video images at the same time.”

'303 Patent, 4:22-42.

115. Put simply, the '303 Patent describes techniques for transmitting encoded video information across communications channels and decoding and reconstructing the video information at a receiver. *See, e.g., id.*, 1:18-21, 1:23-30, 5:6-12, 5:19-25.

4. Brief Description

116. The '303 Patent describes the “transmission structure of the video image” as including “a basic hierarchy and a supplementary hierarchy.” '303 Patent, 3:63-64. According to the '303 Patent, a “basic hierarchy is a hierarchy for transmitting an overview video image, and a supplementary hierarchy is a hierarchy for transmitting an extension video image.” The '303 Patent explains that the basic hierarchical pictures “refer to only basic hierarchical pictures,” which “allows for decoding in the presence of only basic video image coding data.” '303 Patent, 9:61-64. “Meanwhile, allowing supplementary hierarchical pictures to refer to both basic hierarchical pictures and supplementary hierarchical pictures increases the coding efficiency.” *Id.*, 9:64-67. This means that the basic hierarchical pictures of the basic video image coding data represent base layer video data that can be decoded on its own to reproduce a video. Indeed, claim 1 recites “decoding the basic video image coding data so as to reproduce a video image.” '303 Patent, claim 1. The supplementary video image coding data, and the supplementary hierarchical pictures included in the supplementary video image coding data, are therefore used for enhancement purposes and rely on the basic video image coding data for decoding.

117. At the transmitter, the video image coding unit “generates basic video image coding data that corresponds to the basic hierarchy and supplementary video image coding data that corresponds to the supplementary hierarchy.” ’303 Patent, 4:11-14. The transmitter sends the basic video image coding data to the receiver and “stores the supplementary video image coding data in the memory unit 103.” *Id.*, 4:15-18. The receiver “receives the basic video image coding data transmitted” from the transmitter and “stores the received basic video image coding data in a memory of the coding data reconstruction unit 202.” *Id.*, 4:50-51, 5:8-9. The receiver “decodes the basic video image coding data provided from the receiving unit 200 and outputs the decoded basic video image coding data to terminal 2.” *Id.*, 5:14-18; *see also id.*, 7:30-34 (describing generating a “reproduction video image” from the basic video image coding data and outputting the “reproduction video image from the terminal”). At the transmitter, a “memory unit stores supplementary video image coding data of the supplementary hierarchy.” *Id.*, Abstract. The transmitter later transmits the supplementary video image coding data. *E.g., Id.*, 4:22-33, 4:39-42. Put simply, the transmitter first transmits the basic video image coding data, which is stored (and in some case reproduced) at the receiver; and stores the supplementary video image coding data in a memory at the transmitter to be sent later. *Id.*; *see also id.*, 4:40-42.

118. The transmitter “belatedly transmits supplementary video image coding data of the past when there is a margin in the transmission rate.” ’303 Patent, 4:40-42.

The '303 Patent explains that, while the basic video image coding data is “transmitted in real time” by the transmitter, “the supplementary video image coding data is belatedly-transmitted video image coding data of the past stored” at the transmitter. *Id.*, 4:34-42. Therefore, “the supplementary hierarchical picture included in the supplementary video image coding data . . . has a coding order and a display order that precede those of the basic hierarchical picture included in the basic video image coding data transmitted in real time.” *Id.*, 5:25-30.

119. After the transmitter transmits the supplementary video image coding data, the receiver “receives the supplementary video image coding data” from the transmitter. '303 Patent, 4:52-53. The receiver “provides the received supplementary video image coding data to the coding data reconstruction unit 202.” *Id.*, 5:10-12. After the receiver receives the supplementary video image coding data, the receiver acquires from memory the basic video image coding data of the past. The receiver “rearranges the order of the pictures between the basic video image coding of the past stored in the memory and the supplementary video image coding data received” by the receiver “so as to reconstruct video image coding data.” *Id.*, 5:19-25. The coding data reconstruction unit “rearranges the order of pictures between the basic video image coding data of the past stored in the memory and the supplementary video image coding data received by the receiving unit 200 so as to reconstruct video image coding data.” *Id.*, 19-23.

120. Put simply, the '303 Patent describes first sending basic video image coding data with basic hierarchical pictures to a receiver, storing the basic video image coding data in a memory, and in some cases reproducing it. The '303 Patent then describes sending supplementary video image coding data with supplementary hierarchical pictures to the receiver. The receiver reconstructs video image coding data using the basic video image coding data of the past stored in memory with newly received supplementary video image coding data.

121. The '303 Patent explains that “a supplementary hierarchical picture included in the supplementary video image coding data transmitted by the second transmission unit 104 has a coding order and a display order that precede those of a basic hierarchical picture included in the basic video image coding data transmitted in real time by the first transmission unit 102.” '303 Patent, 4:34-39.

122. Because “the supplementary hierarchical picture included in the supplementary video image coding data transmitted from the second transmission unit 104 has a coding order and a display order that precede those of the basic hierarchical picture included in the basic video image coding data transmitted in real time from the first transmission unit 102” the as described previously, the video image coding data cannot be reconstructed in combination with the supplementary video image coding data “if the basic video image coding data of the past is not stored in the memory.” '303 Patent, 5:25-36. This means that, in the '303 Patent, the basic video image coding

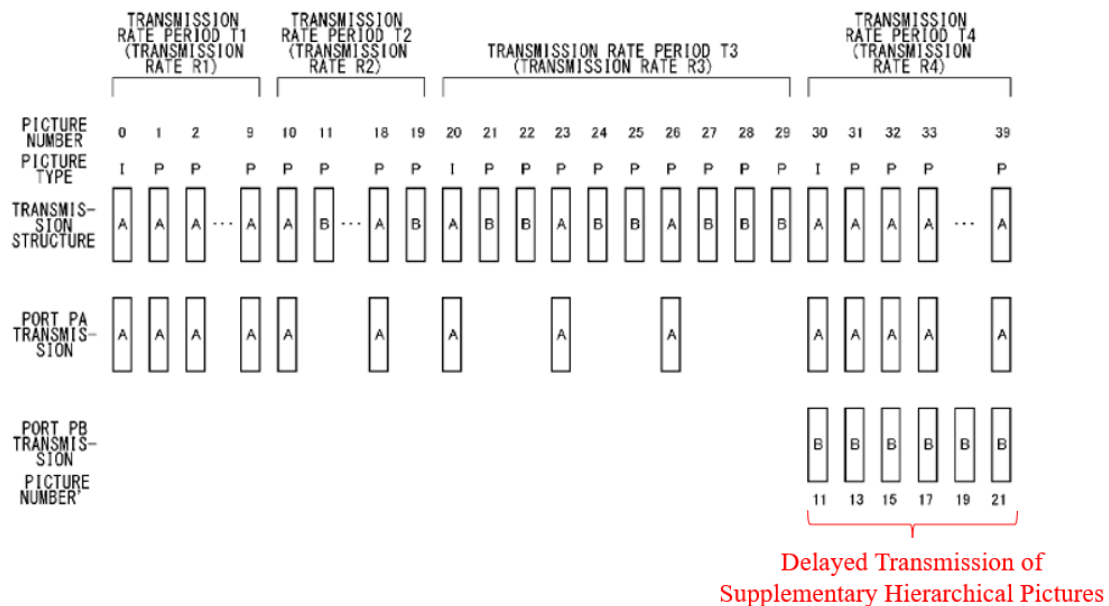
data has been received and is already stored in memory by the time the supplementary video image coding data is received.

123. According to the specification, the transmission of a supplementary hierarchical picture for a given picture or frame in a video sequence can be delayed relative to the transmission of a basic hierarchical picture for the same picture or frame in the video sequence. *See* '303 Patent, 10:25-37. The earlier transmission of a basic hierarchical picture can allow a frame or picture in the video sequence to be decoded and displayed by the receiver earlier (*e.g.*, in real time) at a lower quality when the channel bandwidth is sufficient to support transmission of only the basic hierarchical picture, while the delayed transmission of the basic hierarchical picture can allow the same frame or picture to be decoded and displayed by the receiver at a later time at a higher quality when the channel bandwidth increases by combining (i) the basic hierarchical picture that was earlier stored at the receiver and (ii) the supplementary hierarchical picture that is later received. *Id.*, 10:25-37, 11:38-12:25.

124. For example, as shown in FIG. 5, “[t]he basic hierarchical pictures” (labeled “A”) “are transmitted in real time from the port PA.” ’303 Patent, 11:55-56; *see also id.*, 9:27-36. “When there is a margin in the transmission rate (in a case where the transmission rate is R4), the supplementary hierarchical pictures are transmitted from the port PB in the order stored without exceeding the transmission rate.” *Id.*, 11:56-60. In this manner, “supplementary hierarchical pictures having coding order

and display order that are earlier than those of basic hierarchical pictures transmitted in a given transmission interval are transmitted.”

FIG.5



'303 Patent, FIG. 5 (annotated).

125. The '303 Patent further discloses that “basic video image coding data (basic hierarchical GOP [group of pictures]) and supplementary video image coding data (supplementary hierarchical GOP) are set in units of GOPs” rather than being “set in units of pictures.” '303 Patent, 12:27-36.

5. Comments

126. Nothing I summarized above is new. The prior art teaches first receiving basic video image coding data and storing it. Then, later, receiving supplementary video image coding data, and reconstructing a video using stored basic video image coding

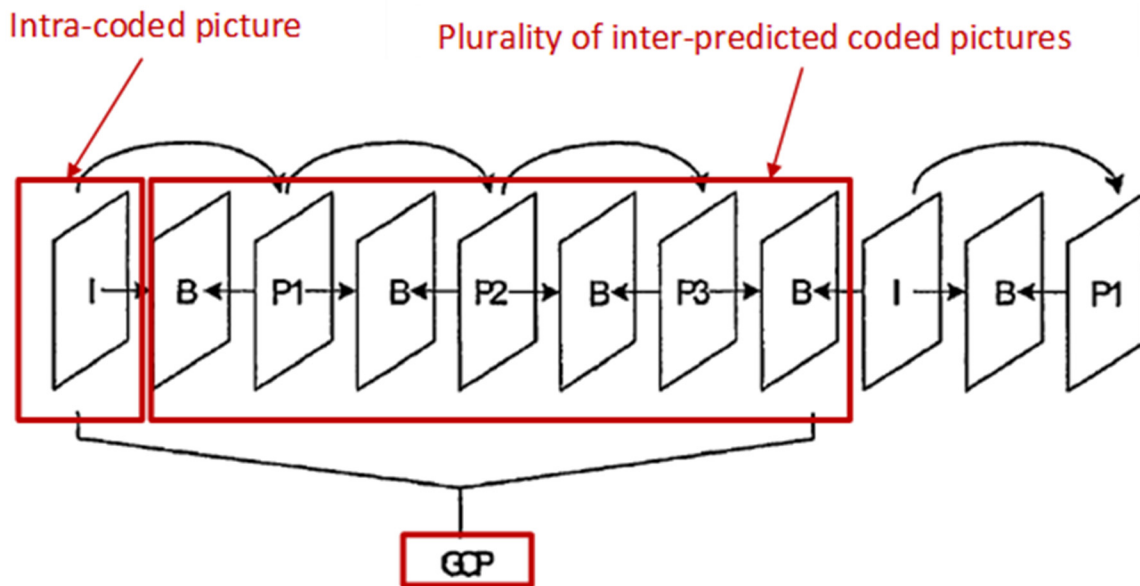
data with received supplementary video coding data. The prior art teaches that it was well known to structure hierarchical video image coding data into GOPs, with each GOP having an intra-coded picture and a plurality of inter-prediction coded pictures.

127. Furthermore, the use of groups of pictures to structure the basic hierarchical pictures and supplementary hierarchical pictures was not new, as evidenced by the prior art I discuss in this declaration, including prior art I use to invalidate the challenged claim and the prior art I discuss as background art. For example, Demircin teaches structuring base layer data and enhancement later data in groups of pictures that include one intra-coded picture and a plurality of inter-prediction coded pictures. *E.g.*, Demircin, FIG. 1. I discuss Demircin at length in the analysis sections of the claims.

128. In another example, Chen, filed in the United States on March 5, 2009, makes it clear that the use of GOPs in hierarchical video coding was commonplace at the time. First, Chen teaches that “known MPEG/H.26x video coding standards (*e.g.* MPEG-1, MPEG-2, MPEG-4, MPEG-4 AVC/H.264, H.263, VC-1), there are basically three types of pictures: I (intraframe coded) pictures, P (interframe coded) pictures and B (bi-directionally predicted) pictures.” Chen, 1:20-24. According to Chen, “[v]ideo sequences are usually coded in a group of picture (GOP) structure.” *Id.*, 1:33-35 (emphasis added).

129. Not only were GOPs commonplace, the “typical MPEG GOP structure” included an “intra encoded picture and more than two predicted encoded pictures.”

E.g., Chen, 1:12-16 (emphasis added). FIG. 1 of Chen illustrates the “typical MPEG GOP structure” that includes an I picture and a plurality of P and B pictures:

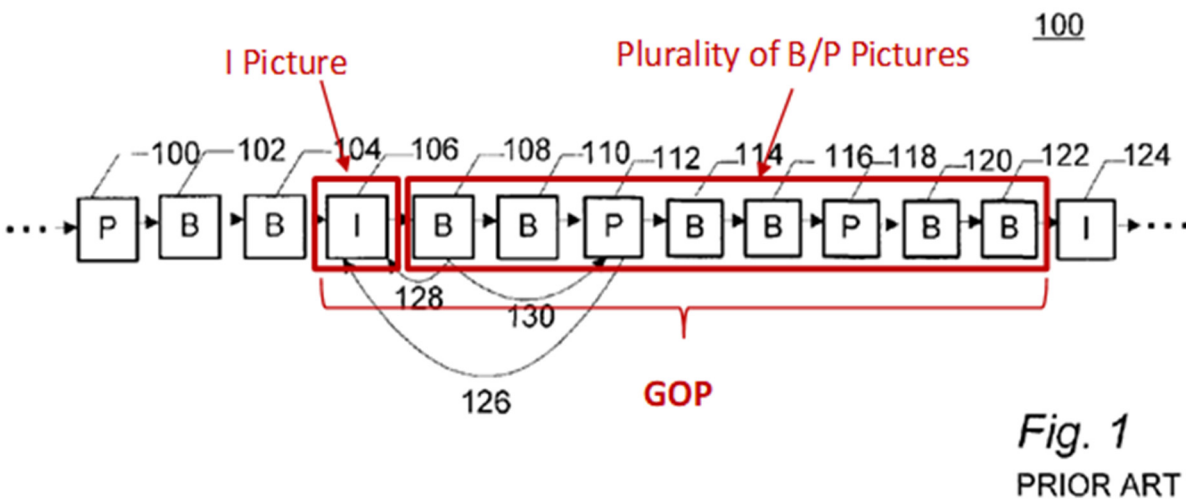


Chen, FIG. 1 (annotated), 4:40. As shown in Chen, it was “typical” in MPEG video coding to structure an I picture and a plurality of B/P pictures into a GOP, as shown above. *Id.*

130. Luthra, filed on June 3, 2003, is another example of prior art that performs video coding using a GOP with an I picture and a plurality of B/P pictures. Luthra teaches, as early as 2003:

Video content typically comprises one or more groups of pictures, each of which are a collection of pictures comprised of one or more macroblocks. Each of the pictures in the group of pictures are either I (intra-coded) pictures, B (bi-directional) pictures, or P (predicted) pictures.

131. Luthra, 1:27-31 (emphasis added). As confirmed by Luthra, an I picture is “intra-coded” and B and P pictures are reference pictures that rely on other pictures for coding and decoding. *Id.*, 1:35-59. Luthra illustrates a prior art group of pictures as including an I picture and a plurality of B pictures and P pictures. *Id.*, FIG. 1:



Luthra, FIG. 1 (annotated). “FIG. 1 is a simplified block diagram illustrating a group of pictures under the MPEG 2 standard.” Luthra, 1:35-38. As described in Luthra, “each individual group of pictures is defined by the collection of pictures beginning at a first I frame and ending at the frame previous to the next I frame.” Luthra, 1:31-34. Thus, in Luthra, the prior art GOP included an intra-coded picture and a plurality of inter-prediction coded pictures.

132. In addition to structuring video coding data into GOPs having I pictures and a plurality of B/P pictures, it was known that the size of the GOP was adjustable. For example, Ansley, filed on June 8, 2011, teaches the ability to adjust the group of

picture size. *E.g.*, Ansley, Abstract. FIG. 2 of Ansley clearly shows that it was known at the time that the size of the group of pictures was not only a well know construct, but that it was also know that the size of the group of pictures could be selected to suit different circumstances or implementation choices.

133. As explained earlier, GOPs are multifunctional and well adapted to the requirements of specific applications. When first conceived, GOPs facilitated trick modes, such as fast-forward and reverse play. Subsequently, for broadcast TV, GOPs facilitate channel surfing. GOPs also provide error resilience for use in error-prone environments.

134. Prior to the priority date of the '303 Patent, there existed numerous products, publications, and patents that implemented or described the functionality claimed in the '303 Patent. Thus, the methodology of the '303 Patent was well-known in the prior art. Further, to the extent there was any problem to be solved in the '303 Patent, it had already been solved in the prior art systems before the priority date of the '303 Patent.

B. Claims

135. The '303 patent includes two claims: the first directed to a video image coding data receiver and the second to a video image coding transmission and reception system. Claims 1 and 2 of the '303 Patent are as follows:

1. A video image coding data receiver comprising:

a processor; and
a memory unit having instructions stored which, when executed by the processor, cause the processor to perform operations comprising:
receiving basic video image coding data;
decoding the received basic video image coding data so as to reproduce a video image;
receiving supplementary video image coding data including a supplementary hierarchical picture whose coding order and display order are earlier by a factor of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures than those of a basic hierarchical picture included in the basic video image coding data, a basic hierarchy and a supplementary hierarchy being set in units of the group of pictures;
acquiring basic video image coding data received before supplementary video image coding data that has been received at the moment; and
reconstructing video image coding data from the basic video image coding data and the supplementary video image coding data.

2. A video image coding data transmission and reception system comprising:
a video image coding data transmitter that codes and transmits a video image; and
a video image coding data receiver that receives and reconstructs video image coding data,
wherein the video image coding data transmitter includes:

a first processor; and
a first memory unit having instructions stored which,
when executed by the first processor, cause the first
processor to perform operations comprising:
acquiring the transmission rate of a network;
setting a transmission structure including a basic hierarchy and a
supplementary hierarchy, the basic hierarchy and the
supplementary hierarchy being set in units of a group of pictures
including an intra coded picture and a plurality of inter prediction
coded pictures;
transmitting basic video image coding data of the basic hierarchy,
wherein the first memory unit is further configured to store
supplementary video image coding data of the supplementary
hierarchy;
transmitting the supplementary video image coding data stored in
the memory unit; and
controlling the transmission of the supplementary video image
coding data according to the transmission rate, and
wherein the video image coding data receiver includes:
a second processor; and
a second memory unit having instructions stored which, when
executed by the second processor, cause the second processor to
perform operations comprising:
receiving the basic video image coding data;
decoding the received basic video image coding data so as to
reproduce a video image;

receiving the supplementary video image coding data including a supplementary hierarchical picture having coding order and display order that are earlier by a factor of the group of pictures than those of a basic hierarchical picture included in the basic video image coding data;

acquiring basic video image coding data received before supplementary video image coding data that is currently received; and

reconstructing video image coding data from the basic video image coding data and the supplementary video image coding data.

C. Prosecution History of the '303 Patent

136. I have reviewed the Prosecution History of the '303 Patent, which issued from U.S. Patent Serial No. 14/657,359, which was filed on March 13, 2015.

137. The application resulting in the '303 Patent was filed on March 13, 2015, by JVC Kenwood Corporation. '303 Patent, Face. The '303 Patent claims the benefit of foreign priority to JP 2014-0071975, dated March 31, 2014. The '303 Patent issued on May 29, 2018. For the purposes of my analysis, I assume that the priority date of March 31, 2014 is the so-called "Critical Date" for the '303 Patent. The prior art I discuss for invalidity and background all predate the Critical Date of the '303 Patent.

138. The Applicant of the 14/657,359 patent application originally submitted 11 claims, 5 of which were independent claims. The claims underwent 4 rejections, each set of rejections compelling the Applicant to make amendments to the claims (or to cancel claims). The '303 Patent includes two independent claims: claim 1 was

originally claim 8 and is generally directed to a video image coding data receiver; claim 2 was originally claim 11 and is generally directed to a video image coding data transmission and reception.

139. During examination of the application, the Examiner first found that Radha (U.S. Pat. 6,292,512) taught both base layer (basic hierarchy) and enhancement layer (supplementary hierarchy). The examiner then rejected claim 8 as obvious in view of Radha and Van Der Schaar (U.S. Pat. 6,697,426). SAMSUNG-1002, 139-140, 146-147, 188-189. Regarding a limitation in claim 8 that recited similar language to that now recited in elements [1.5.1]-[1.5.2], the Examiner noted that Van Der Schaar discloses an enhancement (supplemental) picture E2 having a coding order and display order earlier than a basic picture P3. *Id.*, 139. In response, the Applicant argued that claim 8 required “that the supplementary hierarchical picture whose coding order and display order are earlier than those of a basic hierarchical picture included in the basic video image coding data.” SAMSUNG-1002, 133-134. The Applicant sought to distinguish Van Der Schaar on the basis that it only described “rearranging the transmission order of the enhancement frames ‘to be the same as the display order of the enhancement frames on the decoder side.’” *Id.*

140. The Examiner subsequently rejected claim 8 as obvious based on Radha (U.S. Pat. 6,292,512) in view of Chen (U.S. Pat. 5,886,736). *Id.*, 93-95. In response, the Applicant amended claim 8 to require that the coding order and display order are earlier

“by a factor of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures” than those of a basic hierarchical picture included in the basic video image coding data, “the basic hierarchy and a supplemental hierarchy being set in units of the groups of pictures.” *Id.*, 66. Following this amendment, the Examiner allowed claim 8. *Id.*, 51.

1. Comments

141. Throughout the prosecution of the '303 Patent application, the examiner found that the prior art taught basic hierarchy as a “base layer” and supplemental hierarchy as an “enhancement layer.” *E.g.*, SAMSUNG-1002, 69-70, 103-04, 138-39, 174-75. As discussed below, Demircin also uses the terms “base layer” and “enhancement layer” to refer to the hierarchical pictures. *E.g.*, Demircin [0035]. As evidenced by the examiner’s rejections of the '303 Patent claims and application of the prior art, the examiner concluded, and I agree, that the “basic hierarchy” encompasses a “base layer” and a “supplemental hierarchy” encompasses an “enhancement layer” in a scalable video coding system.

142. Kimoto uses the terms “lower layer” and “upper layer” to describe the hierarchical structure of the content data. *E.g.*, Kimoto [0019]. During prosecution of Kimoto, the examiner interpreted the term “lower layer” to mean “base quality layer” and upper layer to include one or more “quality layers.” *E.g.*, Kimoto File History at NFOA, 5-6. As noted by the cited prior art from Kimoto, a scalable multi-layer

encrypted stream includes “a low quality base layer video (encrypted or non-encrypted)” and “enhancement layer data via different quality layers, each corresponding to a service level.” Zhu, [0026]-[0027]. In addition, the synthesis of lower layer data with upper layer data in Kimoto results in high quality data. *E.g.*, Kimoto [0065]. Thus, the examiner concluded, and I agree, that the lower layer data as base layer data, and the upper layer data as enhancement layer data.

143. As evidenced by the Kimoto File History, the lower layer is the same as a base quality layer, and the upper layer is the same as an enhancement layer. In short, the base layer of Demircin and the lower layer of Kimoto read onto the basic hierarchical coding data of the '303 Patent; and the enhancement layer of Demircin and the upper layer of Kimoto read onto the supplementary video image coding data

144. According to the file history of the '303 Patent, the prior art that I discuss and analyze in this document were not considered by the examiner when rejecting the claims. If Demircin and Kimoto had been considered, as I have explained here, the '303 Patent never should have issued.

VI. Summary of '303 Prior Art

145. The following pieces of prior art render obvious the alleged invention of the '303 patent in the combinations addressed further below (Grounds 1-2).

A. Overview of Demircin

146. Taken together, Demircin and Demircin077² (incorporated by reference into Demircin) disclose a sophisticated, highly detailed system and method for encoding, transmitting, receiving and presenting audiovisual data. The Demircin and Demircin077 references provide statistical methodologies for modeling of transmission channel behaviors and modeling of variable-rate encoded video data together with the perceptual issues related to reproducing encoded, transmitted video data. A key aspect is the requirement for video systems that the data be available for reproduction on a hard deadline.

147. Demircin describes techniques for transmitting coded “audiovisual (AV) data from a source device (*e.g.*, server) to one or more destination devices (*e.g.*, clients) over a packet network.” Demircin, [0023]. Demircin explains that “[a] network linking a video transmitter to a receiver can often be considered a transmission channel with time-varying channel conditions” in which the bandwidth “may vary over time and can be unpredictable” due to various factors. Demircin, [0024]. “Transmission of compressed AV data over such networks is challenging because high-quality AV data streams require a very high bandwidth continuously, and impose stringent delay constraints on delivery of AV data,” and “[f]urthermore, a small number of lost or late

² Demircin077 published as Demircin916 on July 5, 2007.

packets may have a large effect on video quality seen at the receiver, due to the dependencies between parts of the compressed AV stream.” *Id.*

148. Demircin addresses these challenges in the context of Scalable Video Coding (SVC) streams. Demircin, [0027]. “SVC enables coding video in a scalable manner, resulting in a scalable bit stream, for example by coding the video data in several layers.” *Id.*, [0026]. Figure 1 (reproduced below) “illustrates an exemplary prediction structure, GOP structure and layering structure of a sample SVC bit stream” that includes a base layer 10 and enhancement layers 11, 12. *Id.*, [0036]. “[T]he base layer provides a minimum level of video quality,” while the “enhancement layers provide additional quality in an incremental manner, also referred to as progressive refinement.” *Id.*, [0035]. “Decoding of the first enhancement layer requires availability of the base layer.” *Id.* “Decoding of each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” *Id.* The enhancement layers can be used to increase the signal-to-noise ratio (SNR) of coded video content, for example, although Demircin explains that its techniques also apply to other forms of scalability in SVC including temporal scalability (frame rate) and spatial scalability (picture resolution). *Id.*

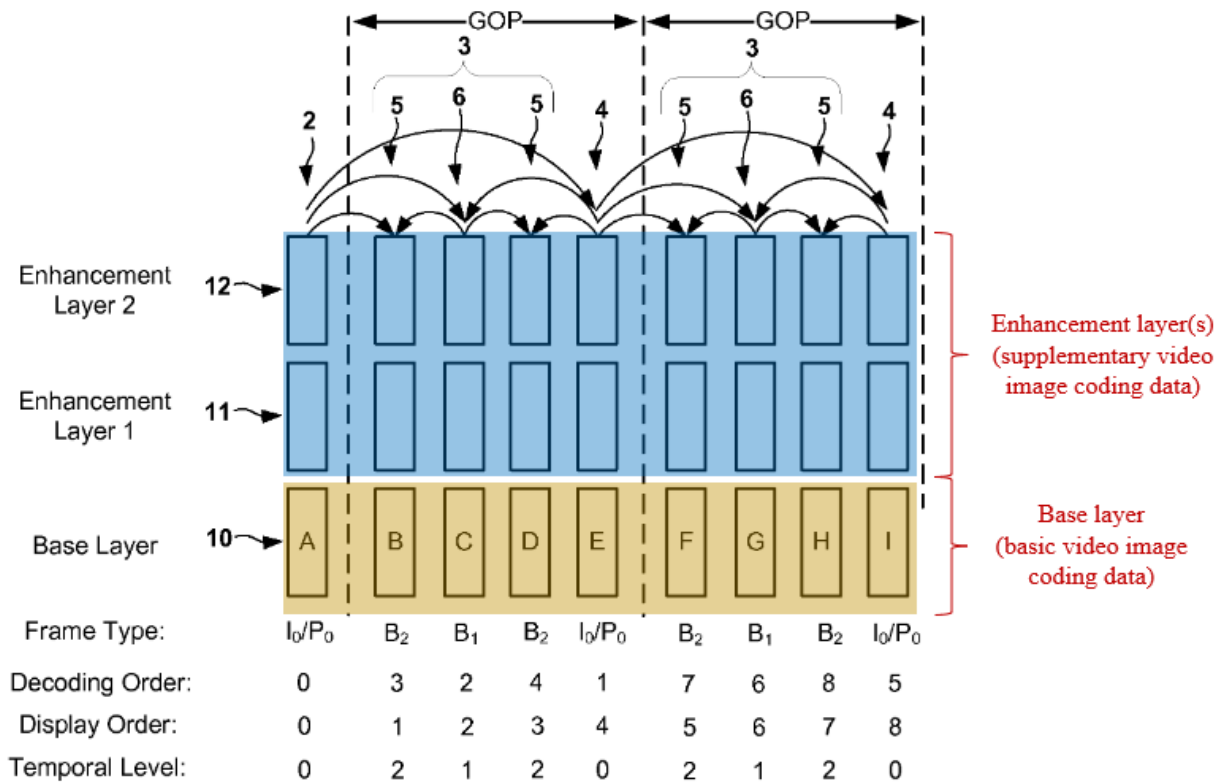


FIG. 1

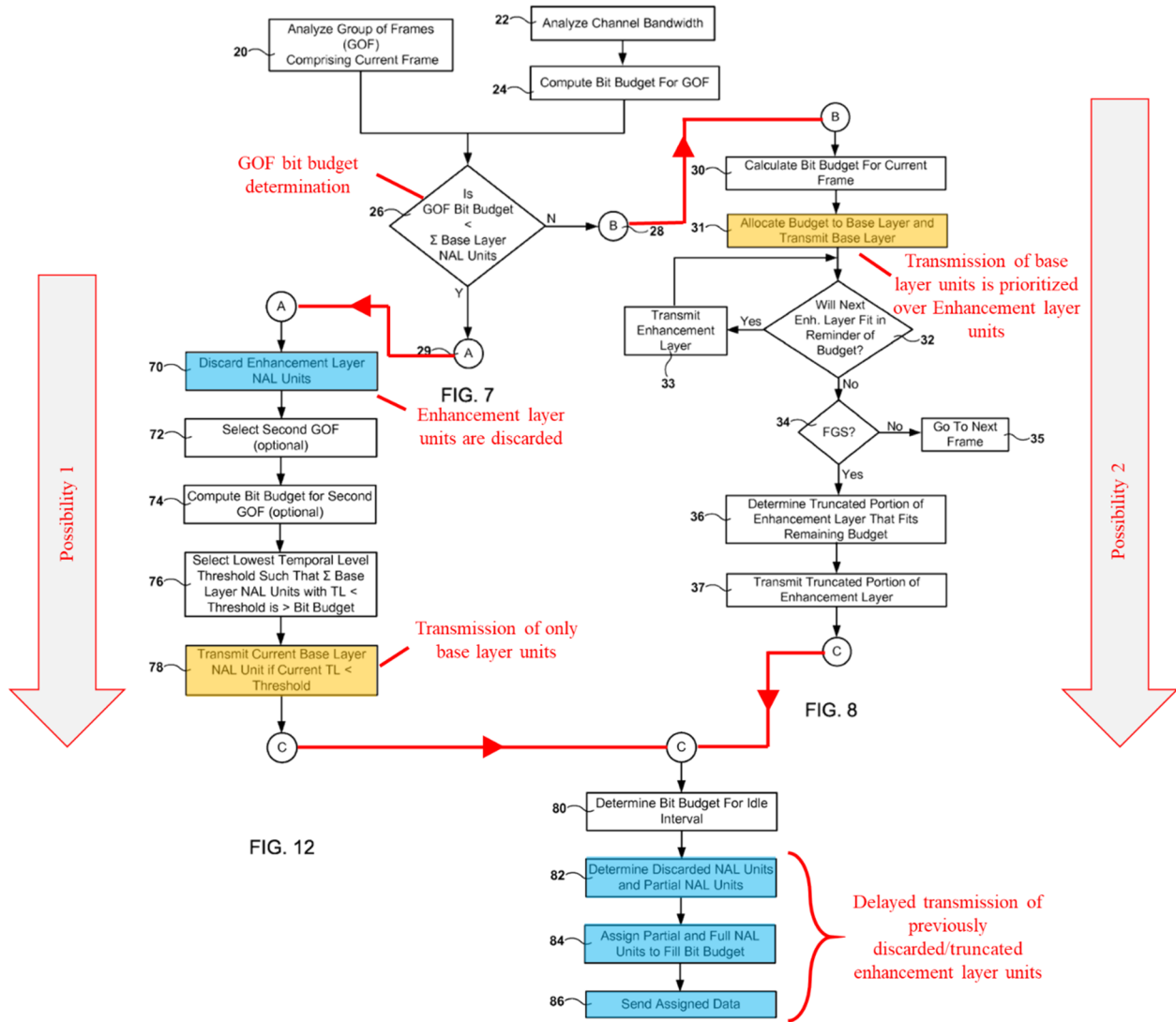
Demircin, FIG. 1 (annotated).

149. Demircin discloses transmitting “units” of data, Demircin077, [0251], that may for example comprise an MPEG GOP (Group Of Pictures). *E.g.*, Demircin, FIG. 1, [0036]; Demircin077, [0309]. For each unit of data, an expected time of delivery may be made for transmitting it so that it arrives before a target delivery time. *See e.g.*, Demircin Figs. 3 & 4, [0040]-[0042]; Demercin077 [0258]. The expectation is made based on factors that include the current channel conditions (including channel bandwidth) and the size of the unit (which can include a “bit budget” Demircin [0040] – the amount of bits to be transmitted). Based on the expectation Demircin decides whether all or only some of the layers for the video will be transmitted, Demircin [0027].

In the latter case, layers that are “discarded” may be held in memory. After transmission of the frames in the current unit starts, Demircin monitors the channel, and if the expectation for channel conditions is incorrect, takes action. If for example there is congestion and the channel bandwidth is reduced, Demircin may discard enhancement layers from video frames, thereby reducing the volume of bits transmitted, and attempt to preserve the delivery deadline. The discarded layers may be held in memory. If on the other hand more than expected bandwidth opens up on the channel, Demircin discloses that the scheduled data will be delivered ahead of the deadline, creating idle time, in which (any) discarded data may now be transmitted.

150. Demircin also discloses starting a new GOP under certain circumstances. *See e.g.*, Demircin, FIG. 12, [0144]-[0145], [0134], [0138]. A scenario in which this could occur is the current bandwidth is limited by congestion, so when planning for a GOP-1, the system runs out of available bits in the bit budget and starts a new GOP-2. That means the current GOP-1 could only have base layer data and makes the current GOP-1 short - say very short: only 4 frames. The enhancement layer data for the current GOP-1 is “discarded” and held in memory. Transmission of this short GOP-1 takes place, then transmission of the new GOP-2 begins but the bandwidth congestion eases and a frame in the new GOP-2 is transmitted quickly, leaving idle time. Transmission of the discarded enhancement GOP-1 data then takes place in the new GOP-2, 5 or more frames later.

151. Demircin's transmission scheme involves "dynamically adapt[ing] a given SVC bit stream for high-quality video streaming, by intelligently selecting which parts of the bit stream should be transmitted, which parts should be discarded entirely, and which parts should be transmitted partially (truncated), in order to meet time-varying rate and/or delay constraints." *Id.*, [0027]. For example, Demircin's system can detect current channel conditions, and based on the current conditions, establishes a "bit budget" for a group of frames (GOF) that is to be transmitted to a remote decoding device (receiver). *Id.*, [0061]-[0062]. The system then selectively transmits all or a portion of the coded data for a GOF that can be accommodated within the allocating bit budget, prioritizing base layer data over enhancement layer data. *Id.*, [0129]-[0131]. If the bit budget cannot accommodate transmission of a frame in the enhancement layer, for example, the enhancement layer frame can be discarded for the moment and later transmitted on a delayed basis during an "idle interval." *Id.*, [0104]-[0122]. The flowcharts in Figures 7, 8, 12, and 13 illustrate these concepts:



Demircin, FIGs, 7, 8, 12, 13 (annotated)

B. Overview of Kimoto

152. Kimoto describes a “moving image data distribution system for distributing reproduction content data that is hierarchically structured so that lower layer data is synthesized with upper layer data to form high quality data.” *E.g.*, Kimoto, Abstract, [0019]-[0024], [0063], [0069]-[0070]. Kimoto acknowledges that with “moving image coding methods defined in MPEG (Moving Picture Experts Group)

and the like, the amount of code to be generated varies greatly from one frame to another.” *Id.*, [0003]. In a viewing situation where bandwidth is limited, “there can occur fluctuations in the reception time of the frame data.” *Id.* Kimoto observed that caching techniques had been proposed “to absorb fluctuations in arrival time of the frames of the first data” received and “to absorb changes in the reproduction time of some of the frame.” *Id.*, [0014]. However, as noted by Kimoto, the “data increases in size as the reproduction contents have a higher bit rate.” *Id.*, [0015]. “This gives rise to the problem that the first data transmission time increases if the first data is transmitted over a limited bandwidth.” *Id.*

153. Thus, Kimoto describes “a moving image data distribution system for distributing moving images” that allows for “quick data initialization even when the reproduction contents have a high bit rate.” Kimoto, [0018]. Kimoto achieves this by “transmitting only the upper layer data when low quality data has already been viewed on a receiving terminal that contains the lower layer data.” *Id.*, [0025]. “When reproducing the contents,” Kimoto describes synthesizing the upper layer data and “the corresponding part 1005 of the lower layer data to generate high quality data.” Kimoto, [0069].

154. Kimoto explains that its techniques are broadly “applicable” to “international standards of the hierarchical coding methods” including “MPEG-2 scalable, MPEG-4 scalable, MPEG-4 FGS, H.264 / MPEG-4 AVC scalable video

coding, and Motion JPEG 2000,” as well as “other hierarchical coding methods.”

Kimoto, [0038]. Kimoto teaches scalability, which involves hierarchical coding:

The transmission data storing unit 100 contains data on reproduction contents. The content data stored is *hierarchically coded*. The data is *scalable* in one or a plurality of directions including time (frame rate), space (resolution), and image quality (bit rate). Some of the data can be extracted for reduced frame rate, resolution, and/or bit rate.

Kimoto, [0037] (emphasis added).

155. To address fluctuations in reception time of video image coding data, Kimoto teaches determining a length of the upper layer data based on transmission bit rate. For example, the “the upper layer first data [has] a data length determined based on a transmission bit rate of the upper layer first data and a reproduction time of the first data of the upper layer.” Kimoto claim 42, [0051]. “The upper layer first data length determining unit 110 calculates a reproduction delay time for the case of transmitting only the upper layer data to the receiving unit 20, and determines an upper layer first data length 1100 (S10). The upper layer first data transmitting unit 101 reads upper layer data as long as the upper layer first data length 1100 from the transmission data storing unit 100, and transmits the upper layer data to the receiving unit 20 (S11).” *Id.*, [0047].

156. To reproduce moving image contents, “the hierarchical data synthesizing unit 203 synthesizes the upper layer first data 1004 stored in above S03 and the corresponding part 1005 of the lower layer data to generate high quality data 1006

(S04).” Kimoto, [0069]. “The reproducing unit 201 then decodes the high-quality data 1006” to reproduce the content at the receiver. *Id.*, [0070] (emphasis added). Accordingly, Kimoto’s receiving unit 20 receives video image coding data in the form of lower layer data and upper layer data from the transmitting unit 10, and reconstructs (i.e., synthesizes) video image coding data.

157. Thus, Kimoto teaches first receiving, at a receiver, lower layer data, storing the lower layer data, and even reproducing the lower layer data before receiving upper layer data. Kimoto, [0041]-[0042]. Later in time, Kimoto teaches receiving and storing the upper layer data at the receiver. *Id.*, [0067]. The, when reproducing the contents, “the hierarchical data synthesizing unit 203 synthesizes the upper layer first data 1004 stored in above S03 and the corresponding part 1005 of the lower layer data to generate high quality data 1006 (S04).” *Id.*, [0069].

VII. Claim Construction

158. I have not been asked to opine on the special meaning of claim terms within the 303 patent. I have been informed that the claim terms should be afforded their plain and ordinary meaning read in the context of the specification and file history of the 303 patent.

159. As I have already mentioned, however, the terms basic video image coding data, basic hierarchical picture, supplementary video image coding data, and supplementary hierarchical picture were implicitly interpreted by the examiner in

mapping art to these terms during original examination of the '303 patent . *Supra*, § V.D.

160. The Examiner's interpretations are consistent with the plain meaning of the terms.

VIII. Ground 1 – Claims 1-2 are Obvious under 35 U.S.C. § 103 based on Demircin

A. Analysis

161. In my opinion, a person of ordinary skill in the art would have found claims 1-2 of the '303 patent to have been obvious based on Demircin.

1. Claim 1

[1.P] A video image coding data receiver comprising:

162. To the extent the preamble is limiting, Demircin discloses a video image coding data receiver in a video image coding transmission and reception system. For example, Demircin discloses “system[s] for adapting a video transmission rate.” Demircin, Abstract. Demircin discloses a destination device (*a video image coding data receiver*) that receives video image coding data from a source (*e.g., a server*) (*a video image coding data transmitter*) over a packet network. Demircin, [0023] (“transmission of audiovisual (AV) data from a source device (*e.g., server*) to *one or more destination devices (e.g., clients)* over a packet network”) (emphasis added), [0024] (“a receiver”), [0119]

(“client station”). The transmitter encodes and transmits video image coding data while the receiver decodes and reconstructs video image coding data. *Infra*, [1.3], [1.5.1], [1.7].

163. Demircin also incorporates by reference U.S. Patent Application Serial No. 11/479,077 (“Demircin077”), which published as U.S. Pub. 2007/0153916 (“Demircin916,” SAMSUNG-1007). Demircin, [0063]. Through Demircin077, Demircin discloses additional details regarding video receivers. For example, with reference to FIG. 1, Demircin077 describes “a plurality of client receiver units 230, 240, 250,” each of which constitutes a video image coding data receiver as recited in [1.P]. *See* Demircin077, [0052], FIG. 1; *see also id.*, [0055], [0195], FIG. 30 (depicting “Receiver”).

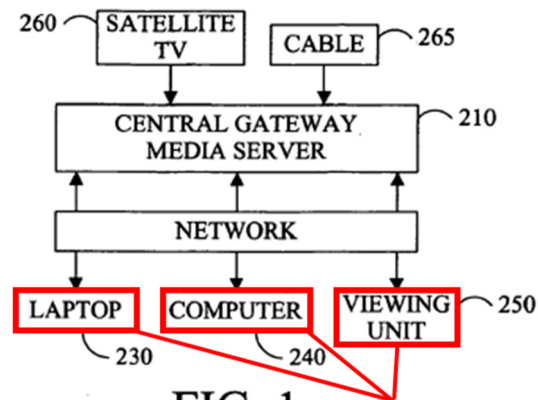


FIG. 1 Video image coding
data receiver

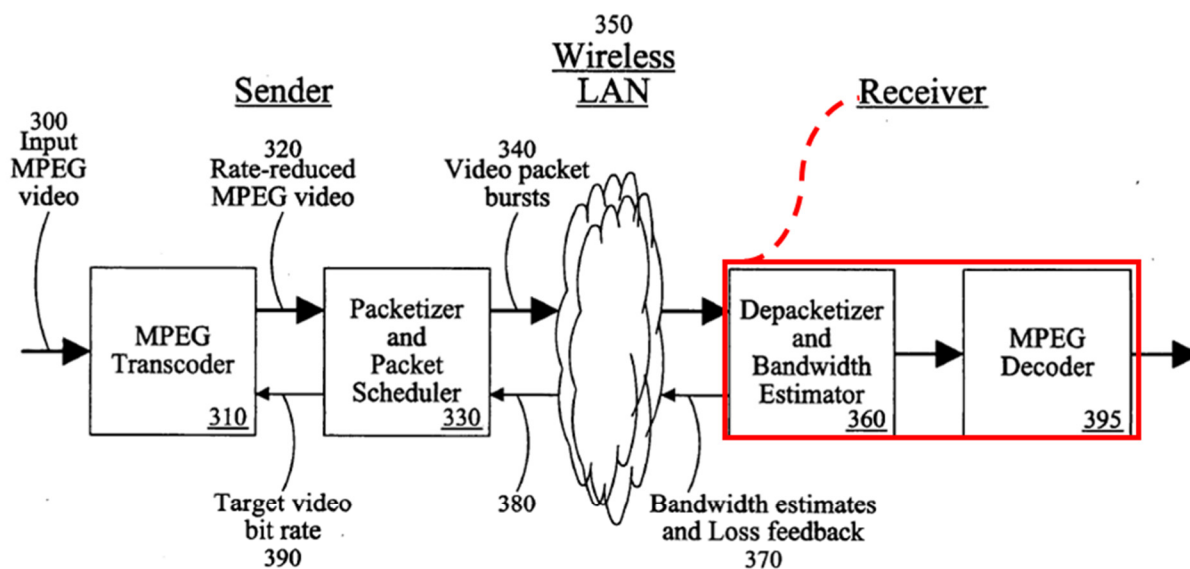


FIG. 30

164. Demircin077, FIGS. 1, 30 (annotated).

[1.1] a processor; and

165. Demircin discloses and renders obvious a “processor” in the video image coding data receiver as recited in [1.1]. For example, Demircin discloses “system[s] for adapting a video transmission rate” that require “computational resources” for processing coded video data. Demircin, Abstract, [0025]-[0026]. Demircin’s “computational resources” would have included a processor for decoding and performing other processing on coded video data at the receiver. Demircin, [0019] (“Elements of embodiments of the present invention may be embodied in hardware, firmware and/or software. While exemplary embodiments revealed herein may only

describe one of these forms, it is to be understood that one skilled in the art would be able to effectuate these elements in any of these forms while resting within the scope of the present invention.”).

166. Demircin further discloses, through Demircin077, which is incorporated by reference in Demircin (*see* SAMSUNG-1005, [0063]), that “[d]ifferent devices interconnected to the network have different resources and different usage paradigms” including “different microprocessors.” Demircin077, [0055] (emphasis added). Therefore, I understand Demircin to teach a processor through the incorporation of Demircin077.

[1.2] a memory unit having instructions stored which, when executed by the processor, cause the processor to perform operations comprising:

167. Demircin renders obvious a memory unit having instructions stored which, when executed by the processor, cause the processor to perform operations. For example, Demircin teaches that “[e]lements of embodiments of the present invention may be embodied in hardware, firmware and/or software,” and a memory unit as recited in [1.2] was commonly and ordinarily included in “hardware, firmware and/or software” implementations of video image coding data receivers. Demircin, [0019].

168. Demircin further discloses through Demircin077916, which is incorporated by reference in Demircin (*see* SAMSUNG-1006, [0063]), that “[d]ifferent devices interconnected to the network have different resources and different usage

paradigms” including “different memory requirements.” Demircin077, [0055] (emphasis added). For these reasons, I understand Demircin’s video image coding data receiver to include a memory unit having instructions as recited in [1.2] to store instructions for carrying out the disclosed processing functions of the receiver.

[1.3] receiving basic video image coding data;

169. Demircin renders obvious a receiver that receives basic video image coding data. For example, Demircin discloses that the receiver receives network abstraction layer (NAL) units for base layer coding data (*basic video image coding data*), which were transmitted by a source/server (*video image coding data transmitter*). *Id.*

170. In more detail, Demircin describes “Scalable Video Coding (SVC),” which “is an extension of the H.264/MPEG AVC video codec.” Demircin, [0035]. “SVC supports scalability in the spatial, temporal and SNR (signal-to-noise ratio) dimensions.” *Id.* “A base layer provides a minimum level of video quality, and may provide compatibility with H.264/AVC.” *Id.* “One or more enhancement layers provide additional quality in an incremental manner, also referred to as progressive refinement.” *Id.* “Decoding of the first enhancement layer requires availability of the base layer” and “[d]ecoding of each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” *Id.* “The SVC bitstream comprises NAL (Network Abstraction Layer) units similar to H.264/AVC.” *Id.*, [0038]. “The base layer data of a video frame may be encapsulated into a NAL unit[]” *Id.*

171. Through FIG. 1 (reproduced below), Demircin “illustrates an exemplary prediction structure, GOP structure and layering structure of a sample SVC bit stream.” Demircin, [0036]. As shown in FIG. 1, Demircin teaches “a layered structure for each frame” that includes “a base layer 10 and two enhancement/refinement layers 11, 12.” Demircin, [0036]-[0037]. The coded data in base layer 10 for one or more frames (pictures) corresponds to “*basic video image coding data*” as recited in [1.3]. ’303 Patent, 6:24-31. Like the basic hierarchical pictures described in the ’303 Patent, which are coded without reference to any pictures in the supplementary hierarchy, the pictures in Demircin’s base layer 10 are coded without reference to any pictures in enhancement layers 11, 12. Demircin, [0035]; ’303 Patent, 9:57-67.

172. Regarding temporal scalability, the pictures at temporal level 0 in Demircin’s base layer 10 are I-pictures that can be coded without reference even to other pictures at higher temporal levels in base layer 10. *Id.*; ’303 Patent, 11:22-36. The base layer 10 data that is transmitted from the transmitter is received by the receiver. Demircin, [0024], [0061] (“received by the client”), [0062], [0063] (“deliver frames to the client”), [0119].

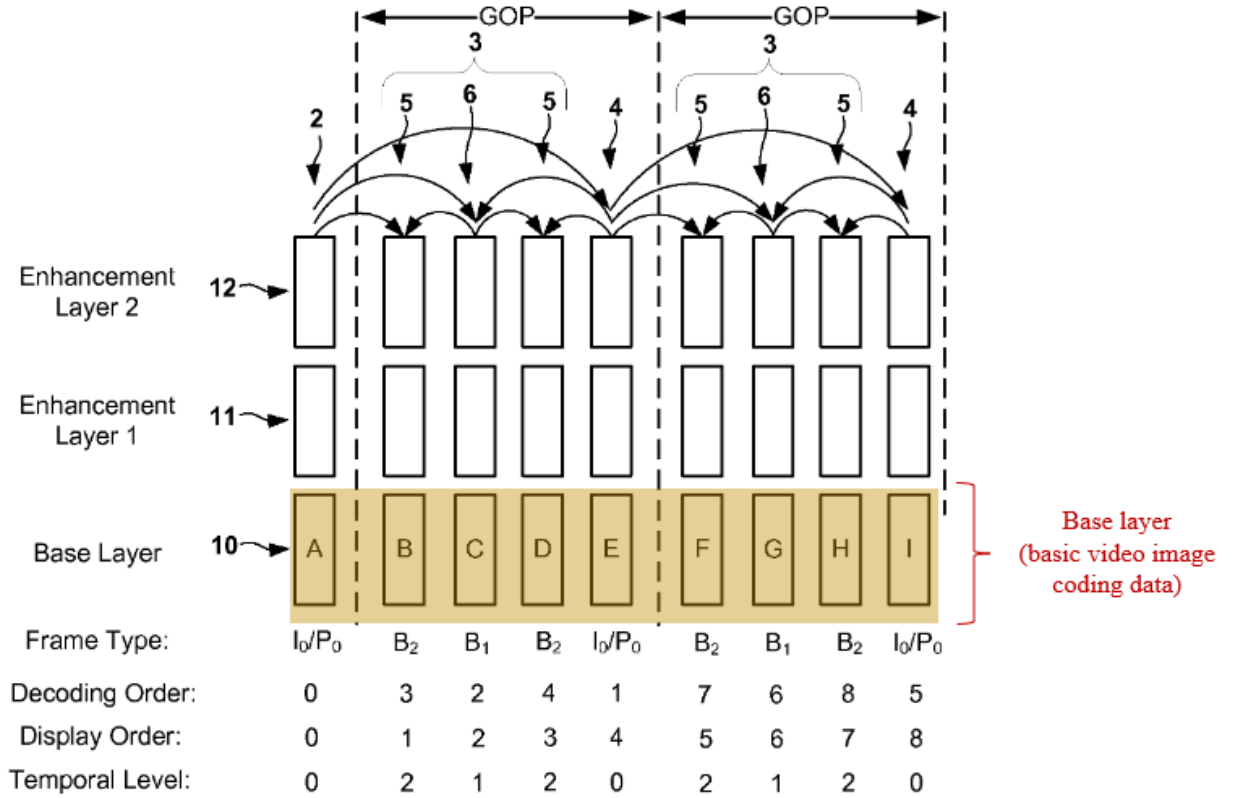
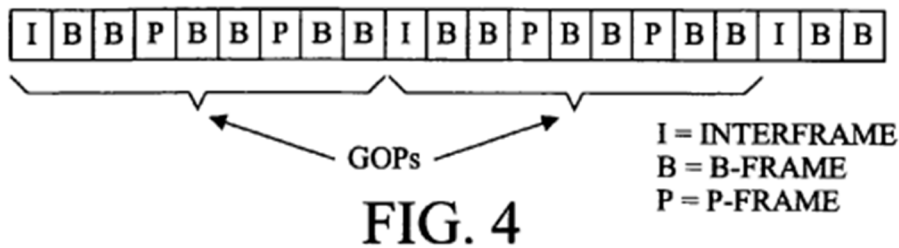


FIG. 1

Demircin, FIG. 1 (annotated).

173. While the GOP size depicted in FIG. 1 accommodates 4 frames, Demircin broadly contemplates other GOP sizes and structures as well. For example, Demircin further discloses through Demircin077, which is incorporated by reference (*see* Demircin, [0063]), an arbitrarily sized GOP of length N_{GOP} (*e.g.*, 9 frames): “The length of a GOP in number of pictures is denoted by N_{GOP} ... in FIG. 4 where $N_{\text{GOP}}=9$ for illustration purposes.” Demircin077, [0069], FIG. 4.



Demircin077, FIG. 4.

[1.4] decoding the received basic video image coding data so as to reproduce a video image;

174. Demircin renders obvious a receiver that decodes the received basic video image coding data so as to reproduce a video image. In Demircin, in situations where only base layer frames are transmitted, the receiver only decodes and displays base layer frames to reproduce an image. Such a situation arises when, for example, the GOF bit budget can only accommodate the transmission of base layer frames. If this claim element is interpreted more broadly such that decoding both base layer and enhancement layer frames satisfies the claim element, then Demircin teaches that, too, when it transmits both base layer and enhancement layer frames within the decoding deadline. In that case, when both the base layer and enhancement layer frames are transmitted within the decoding deadline, both base layer and enhancement layer frames are decoded at the receiver for reproduction of video.

175. As discussed above (*supra*, [1.3]), Demircin discloses that video data can be coded in a “layered structure” in which the “base layer provides a minimum level of video quality” and “[o]ne or more enhancement layers provide additional quality.”

Demircin, [0035], FIG. 1. As described in further detail below, and depending on current network conditions, Demircin's transmitter will transmit, and the receiver will receive, only the base layer NAL units for one or more video frames (basic video image coding data) to the exclusion of enhancement layer NAL units for those frames by the respective decoding and presentation deadlines for those frames. *E.g.*, Demircin, [0035], [0129]-[0132], [0148], FIGS. 7-8, 12, 13. In such instances, the receiver would reproduce the frames with a "minimum level of video quality" by decoding only the base layer NAL units for those frames without the corresponding enhancement layer NAL units for those frames. *Id.*

176. For example, Demircin discloses that the transmitter transmits base layer NAL units to the receiver, and discards enhancement layer NAL units, when the "bit budget" for individual frames or the bit budget for a "group of frames" (GOF) is less than required to transmit both the enhancement layer NAL units and the base layer NAL units:

In these embodiments, once it has been determined that a base layer fits within the bit budget for a GOF, a bit budget for a current target frame may be determined 30. The base layer may then be allocated 31 to the current frame bit budget leaving a current frame budget remainder. The base layer may be transmitted at this time if it has not already been transmitted or it may be selected for transmission at some later time. In some embodiments, the size of the next most important enhancement layer may then be compared 32 to the current frame budget remainder.

... In some embodiments, when the next most important enhancement layer will not fit in the updated budget remainder, it may be determined whether portions of an enhancement layer may be transmitted. ... If a layer is not divisible (*e.g.*, not FGS) the remaining enhancement layer may be discarded or otherwise omitted from the current transmission schedule 35 and the process may proceed to the next frame.

Demircin, [0131]-[0132] (emphasis added); *see also id.*, (“When all base layers do not fit in the GOF bit budget, all enhancement layer units may be discarded or omitted from transmission at the current time 70.”).

177. Demircin teaches that the system should further abide by “decoding deadlines,” which refers to the time by which the decoder must decode a frame for reproduction. Demircin, [0105] (“Embodiments comprising delayed transmission may detect idle channel intervals (instances when the channel is not fully utilized), and subsequently transmit previously dropped or truncated NAL units whose decoding deadlines have not yet expired.”); [0063] (“Some embodiments may deliver frames to the client prior to their decoding deadline, while minimizing the distortion due to rate adaptation.”), FIGS. 7-8, 12. The receiver (decoding device) would only decode base layer data for a frame, without combining such base layer data with enhancement layer data, if the receiver has received only base layer NAL units for the frame by the frame’s decoding deadline. *E.g.*, Demircin, [0035], [0129]-[0132], [0148], FIGS. 7-8, 12, 13. Indeed, this is evident from Demircin’s explanation that “dropped or truncated NAL units” for enhancement layer frames will only be transmitted if their frames’ “decoding

deadlines have not yet expired.” Demircin, [0105]. Demircin thus discloses and renders obvious the scenario where coded base layer data (basic video image coding data) is decoded, even without use of an enhancement layer data. Demircin [0027]-[0032], [0035]. The real-time nature of Demircin implies that the decoded video data would be reproduced after the decoding deadline passes. Thus, Demircin teaches decoding the base layer data to reproduce a video image.

178. To the extent the breadth of [1.4] allows for decoding the received basic video image coding data together with supplementary video image coding data so as to reproduce a video image, Demircin further discloses and renders obvious such features as well. For instance, Demircin teaches that “[o]ne or more enhancement layers provide additional quality in an incremental manner, also referred to as progressive refinement,” “[d]ecoding of the first enhancement layer requires availability of the base layer,” and “[d]ecoding of each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” Demircin, [0035]. When sufficient bit budget exists for both the base layer and enhancement layer NAL units for a frame (video image) to be transmitted by the transmitter and received by the receiver before the frame’s decoding deadline expires, the receiver would decode the base layer and enhancement layer NAL units to reproduce the frame at a higher quality than if only the base layer NAL units were decoded. Demircin, [0131] (“If the budget remainder is at least as large as the next most important enhancement layer, that enhancement layer

may be allocated to the budget thereby creating an updated budget remainder and the enhancement layer may be transmitted 33. This process may be repeated until all enhancement layers for the current frame have been transmitted or until the next most important enhancement layer is larger than the updated budget reminder.”), FIGS. 7-8.

[1.5.1] receiving supplementary video image coding data including a supplementary hierarchical picture . . .

179. Demircin discloses and renders obvious receiving supplementary video image coding data including a supplementary hierarchical picture. *Supra* A.2-[1.3]. For example, Demircin discloses that the receiver receives network abstract layer (NAL) units for enhancement layer coding data (supplementary video image coding data), including enhancement layer NAL unit(s) for an individual frame or picture (supplementary hierarchical picture). *Id.*

180. In more detail, Demircin discloses embodiments involving “Scalable Video Coding (SVC),” which “is an extension of the H.264/MPEG AVC video codec.” Demircin, [0035]. “SVC supports scalability in the spatial, temporal and SNR (signal-to-noise ratio) dimensions.” *Id.* “A base layer provides a minimum level of video quality, and may provide compatibility with H.264/AVC.” *Id.* “One or more enhancement layers provide additional quality in an incremental manner, also referred to as progressive refinement.” *Id.* “Decoding of the first enhancement layer requires availability of the base layer” and “[d]ecoding of each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” *Id.*

“The SVC bitstream comprises NAL (Network Abstraction Layer) units similar to H.264/AVC.” *Id.*, [0038]. “[T]he data for each enhancement layer of a video frame may be encapsulated into a NAL unit.” *Id.*

181. Through FIG. 1 (reproduced below), Demircin “illustrates an exemplary prediction structure, GOP structure and layering structure of a sample SVC bit stream.” Demircin, [0036]. As shown in FIG. 1, Demircin teaches “a layered structure for each frame” that includes “a base layer 10 and two enhancement/refinement layers 11, 12.” Demircin, [0036]-[0037]. The coded data in either or both of enhancement layers 11, 12 for one or more frames (pictures) corresponds to “supplementary video image coding data” as recited in [1.5.1], and includes enhancement layer NAL unit(s) for a “supplementary hierarchical picture” for a given frame in the video image sequence. ’303 Patent, 6:24-31. Like the supplementary hierarchical pictures described in the ’303 Patent, which can be coded with reference to pictures in both the basic and supplementary hierarchies, the pictures in Demircin’s enhancement layers 11, 12 can also be coded with reference to both pictures in the base layer 10 and other pictures in the enhancement layers 11, 12. Demircin, [0035]; ’303 Patent, 9:57-67. Relevant to temporal scalability, the pictures at temporal levels 1 and 2 in Demircin’s both base layer 10 and enhancement layers 11, 12 are bidirectionally inter-prediction with reference to other pictures at lower temporal levels. *Id.*; ’303 Patent, 11:22-36. Therefore, the pictures at temporal levels 1 and 2 in base layer 10 can alternatively be

characterized as “supplementary video image coding data including a supplementary hierarchical picture” as recited in claim 1 when only the pictures at temporal level 0 in base layer 10 are characterized as “basic video image coding data” including a basic hierarchical picture. *Id.*

182. Data from enhancement layers 11, 12 that is transmitted from the transmitter is received by the receiver. Demircin, [0024], [0061] (“received by the client”), [0062], [0063] (“deliver frames to the client”), [0119]. Demircin’s transmitter “transmits enhancement layer NAL units of the current frame provided that the enhancement NAL units of other frames at the same layer fit into the bit budget for the group,” for example. Demircin, [0030].

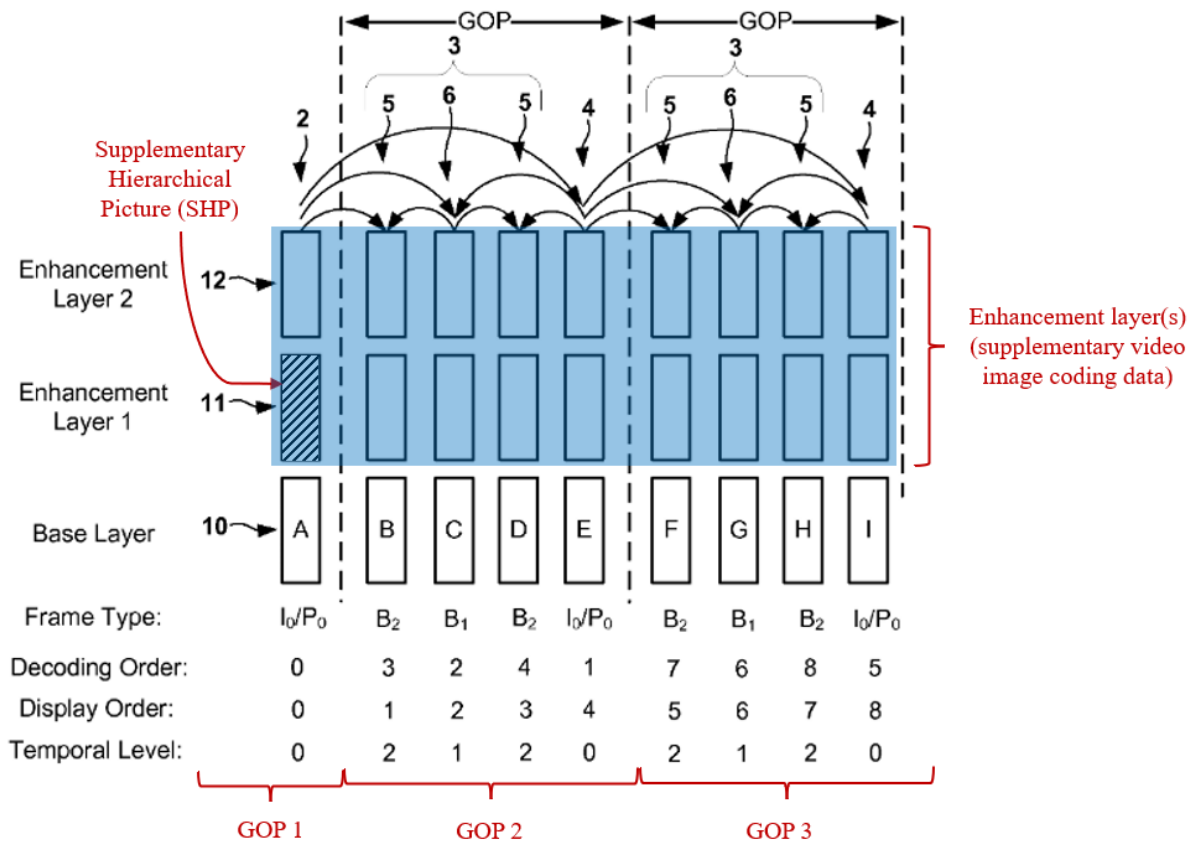


FIG. 1

Demircin, FIG. 1 (annotated).

[1.5.2] [a supplementary hierarchical picture] whose coding order and display order are earlier by a factor of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures than those of a basic hierarchical picture included in the basic video image coding data,

183. Demircin discloses and renders obvious the features recited in [1.5.2].

184. For example, in FIG. 1, Demircin “illustrates an exemplary prediction structure, GOP structure and layering structure of a sample SVC bit stream” that “uses I-frames 2, which are intra-predicted from within the frame; B-frames 3, which may be predicted from one or more other frames,” and optionally “P-frames 4 which are

predicted from another I-frame.” Demircin, [0036]; *see also id.*, [0039], [0043], [0095]. The I-frame within Demircin’s GOP is an “intra coded picture” as recited in [1.5.2] and the series of three B-frames in the GOP constitutes “a plurality of inter prediction coded pictures” as recited in [1.5.2]. *Id.*, [0036] (“A hierarchical B-frame 5 may be predicted from other B-frames as well as I-frames and P-frames while a standard B-frame 6 may only be predicted from I-frames and P-frames.”). Demircin’s description of I-, P, and B- frames is consistent with the description of intra-coded and inter-prediction coded pictures in the specification of the ’303 Patent itself. ’303 Patent, 9:19-23 (“Picture types include an I picture for performing intra coding, a P picture for performing inter prediction coding, and a B picture for performing inter prediction coding in a bidirectional manner.”).

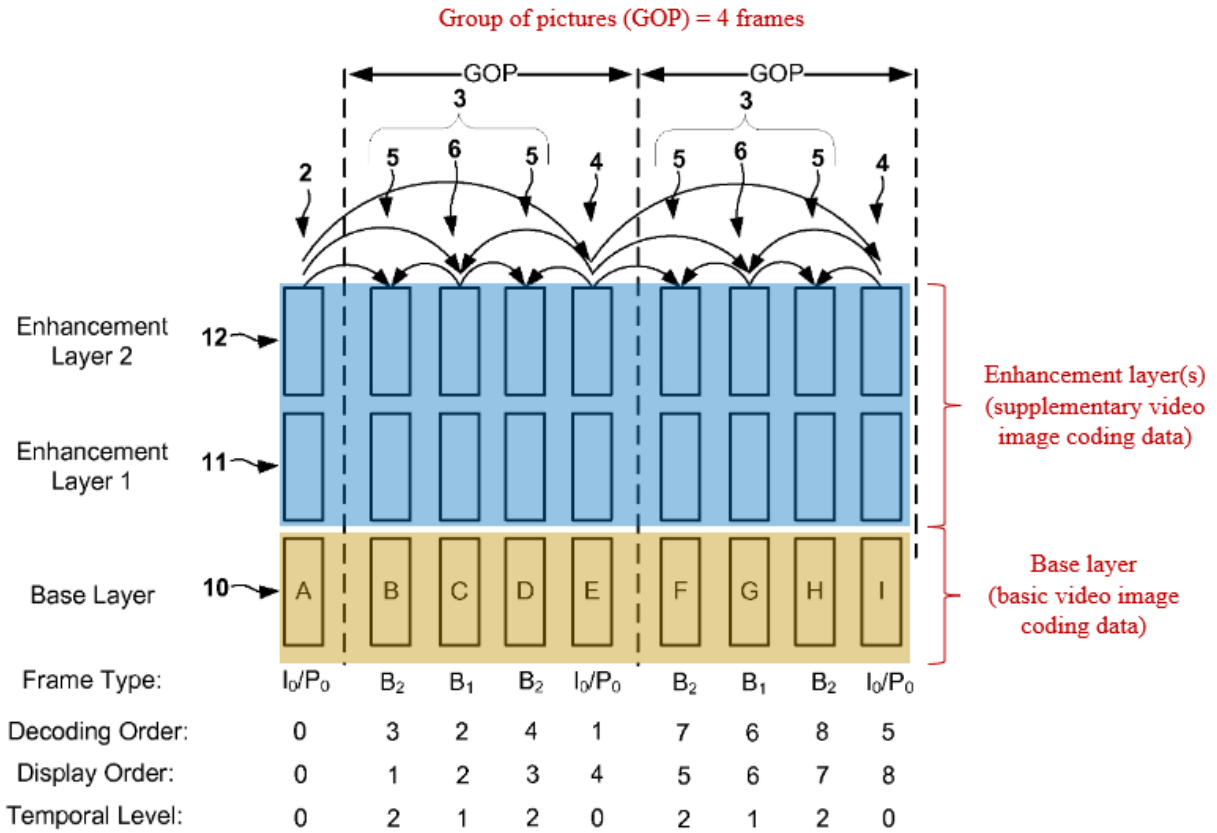


FIG. 1

Demircin, FIG. 1 (annotated).

185. Demircin also discloses and renders obvious a supplementary hierarchical picture whose coding order and display order are earlier by a factor of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures than those of a basic hierarchical picture included in the basic video image coding data. Consider the following illustration from FIG. 1 of Demircin, for example:

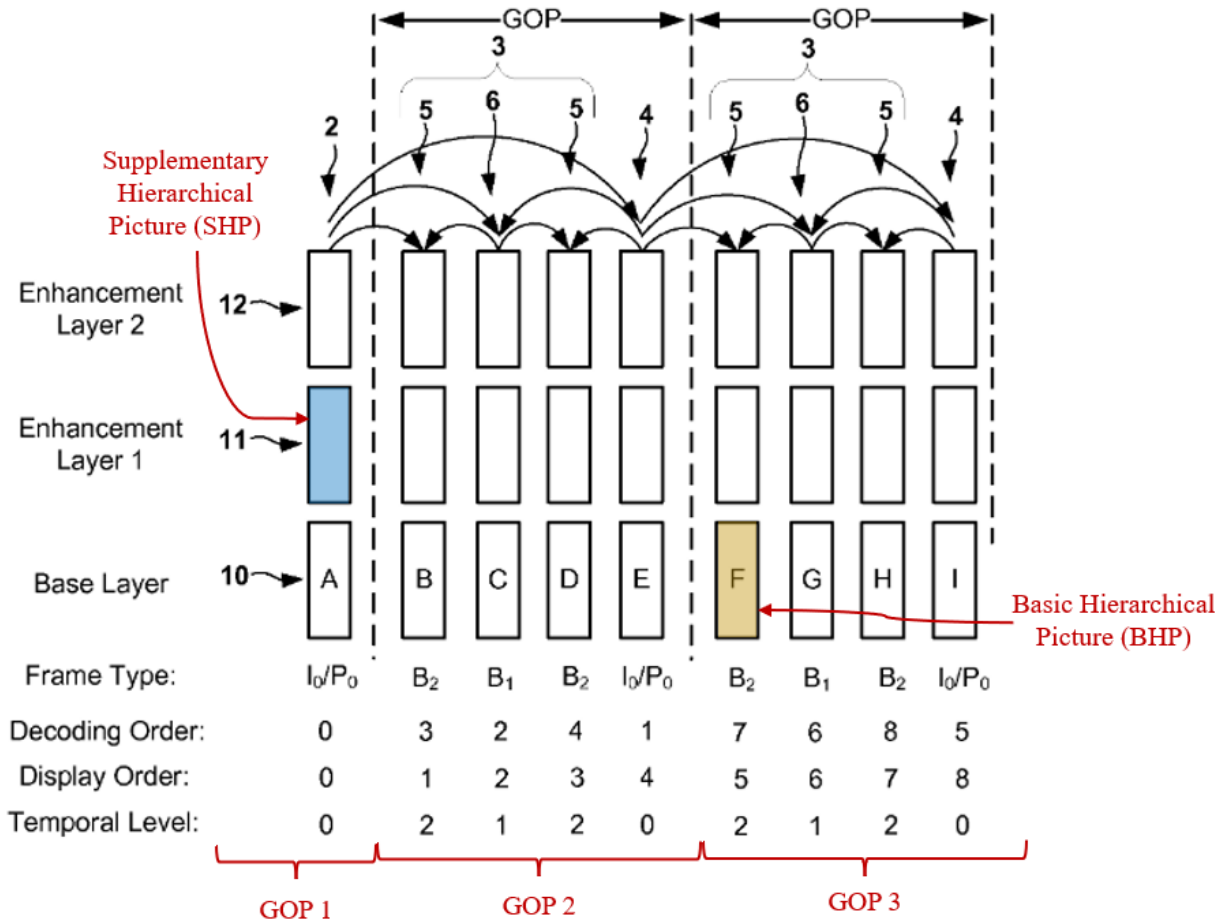


FIG. 1

Demircin, FIG. 1 (annotated). As shown above, Demircin's prediction, GOP, and layering structure includes an enhancement layer picture for frame 0 in display order (supplementary hierarchical picture ("SHP")) indicated by blue shading and a base layer picture for frame 5 in display order (basic hierarchical picture ("BHP")) indicated by gold shading.³ The blue-shaded SHP in this example has a display order that is earlier

³ The shaded pictures in this figure are merely examples provided to illustrate one scenario in which Demircin discloses and renders obvious [1.5.2]. Similar relationships exist between other base layer

by a factor of a group of pictures than the gold BHP, which is evident from the display order annotations in FIG. 1. Specifically, the blue-shaded SHP in the enhancement layer 11 has a display order index of “0” while the gold-shaded BHP in the base layer 10 has a display order index of “5.” The GOP size in FIG. 1 is just four frames, and thus the display order of the blue-shaded SHP is earlier by a factor of a group of pictures including an intra-coded picture and a plurality of inter-prediction coded pictures than the display order of the gold-shaded BHP. *Id.*

186. Likewise, the blue-shaded SHP in this example has a coding order that is earlier by a factor of a group of pictures than the gold-shaded BHP, which is evident from the decoding order annotations in FIG. 1. The decoding order indicates the order in which frames are decoded at the receiver and is both a type of “coding order” itself, and is the same as the order in which frames are encoded at the source. Here, the blue-shaded SHP has a decoding order index of “0” while the gold-shaded BHP has a decoding order index of “7.” The GOP size in FIG. 1 is just four frames, and thus the decoding order of the blue-shaded SHP is earlier by a factor of a group of pictures including an intra-coded picture and a plurality of inter-prediction coded pictures than the decoding order of the gold-shaded BHP. *Id.*

and enhancement layer pictures in Demircin’s structure, and further, the enhancement layer picture (SHP) could be identified from either or both of the enhancement layers 11, 12.

187. Although the specification of the '303 Patent describes certain embodiments in which the transmission of a supplementary hierarchical picture for a frame is delayed by a period of time (*e.g.*, multiple transmission rate periods) relative to transmission of a basic hierarchical picture (*see, e.g.*, '303 Patent, 11:38-12:25 and FIG. 5), claim 1 recites no language that requires the “supplementary hierarchical picture” (SHP) in [1.5.2] to be received after the “basic hierarchical picture” (BHP) in [1.5.2]. Nonetheless, Demircin discloses and renders obvious [1.5.2] and the features of claim 1 as a whole regardless of whether the receiver must receive the SHP after the BHP. *Id.* This follows because Demircin discloses both scenarios in which (1) an enhancement layer NAL unit for a frame (SHP) is transmitted, and hence received, without delay during the ordinary time of transmission/reception for the GOP to which the frame belongs, as well as scenarios in which (2) the enhancement layer NAL unit for a frame (SHP) is transmitted, and hence received, on a delayed basis by well over a factor of a GOP.

188. *Demircin's First Example Scenario: Non-Delayed Transmission of Enhancement Layer NAL Unit(s)*

189. Considering the first example scenario in Demircin in which an enhancement layer NAL unit for a frame (SHP) is transmitted, and hence received, without delay during the ordinary time of transmission/reception for the GOP to which the frame belongs, it is clear that the receiver would receive the blue-shaded SHP

identified in the annotated version of FIG. 1 (above) before the gold-shaded BHP.⁴ For example, Demircin teaches that frames are grouped into groups of frames (GOFs) for purposes of computing “bit budgets” and setting transmission parameters.⁵ Demircin, [0029] (“[A] group of frames comprising a current frame and several subsequent frames is considered, a bit budget for that group of frames is calculated, and then bits are allocated to the Network Abstraction Layer (NAL) units of the current frame according to several methods. ... The use of a moving window comprising a

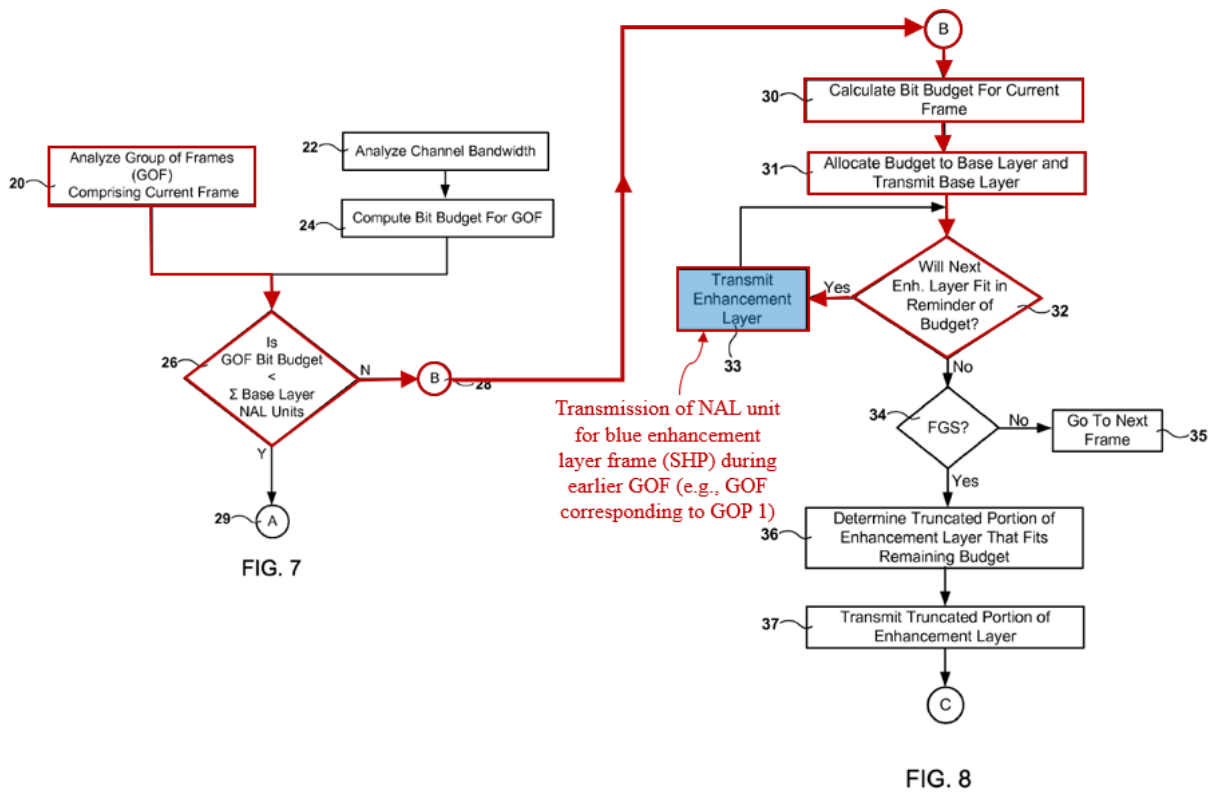
⁴ Notably, regardless of the relative order of transmission and reception between the blue-shaded SHP and the gold-shaded BHP, the display order and coding order of the blue-shaded SHP always remains earlier by at least a factor of a group of pictures including an intra-coded picture and a plurality of inter-prediction coded pictures than the display order and coding order of the gold base layer frame (BHP).

⁵ The MPEG standard defines “picture” to be a “collective term for a field or a frame.” SAMSUNG-1015, 8. Thus, a GOF and GOP can be taken to mean the same thing.

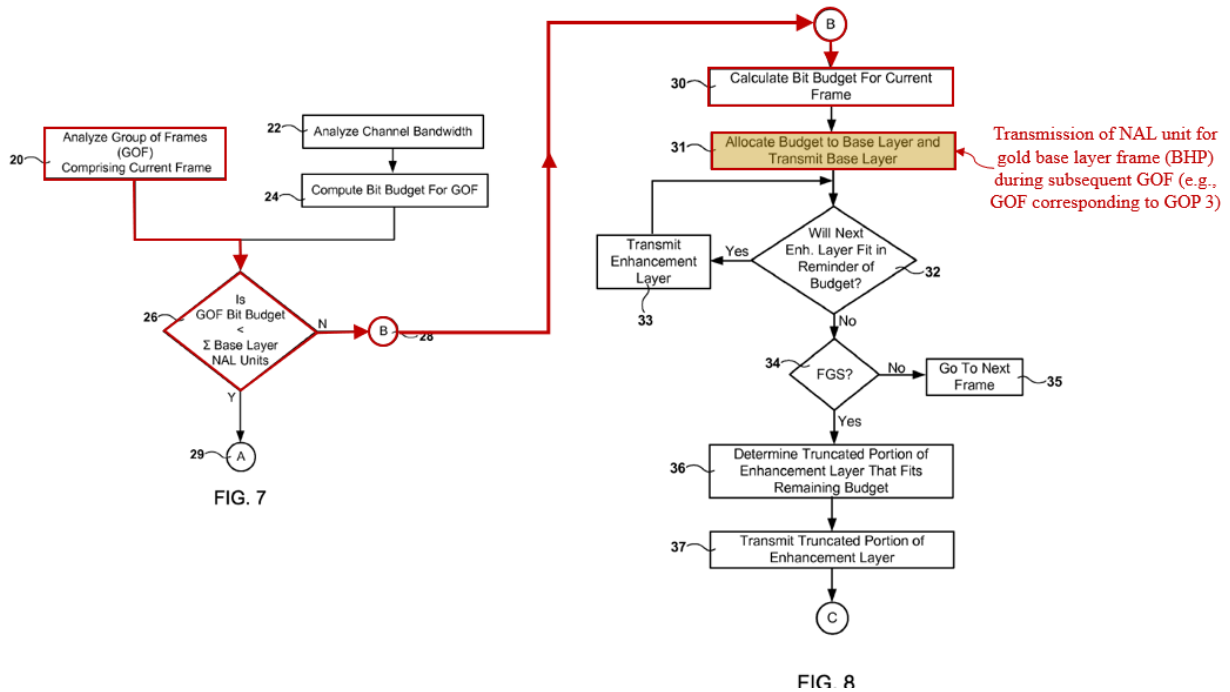
To the extent Demircin does not expressly disclose alignment between GOFs and GOPs, it would have been obvious to implement GOFs with frames that are coextensive with the grouping of frames for GOPs. Indeed, Demircin confirms that GOPs and GOFs comprise a same or similar number of frames (e.g., “a current frame and several subsequent frames” for a GOF and “four” frames for a GOP). Demircin, [0029], [0036], FIG. 1. The selection of a GOF length is arbitrary and flexible, and Demircin leaves the selection of GOF length to be based on performance considerations. *See id.*, [0099] (describing different GOF sizes that may “result in better performance”). While the mappings in this ground do not require alignment or co-extensiveness between GOFs and GOPs, for simplicity of explanation and ease of illustration, Petitioner nonetheless treats the GOFs as consisting of the same groupings of frames as the GOPs. *Id.*

group of multiple frames reduces video quality variations over time due to the variations in the bit sizes of individual frames.”); *see also id.*, [0099], [0128].

190. Following the path through steps 20, 26, 28, 30, 31, 32, and 33 in FIGS. 7-8, Demircin discloses how an enhancement layer picture (SHP) for a frame (such as that indicated by the blue shading in the annotated version of FIG. 1) can be transmitted without delay during an initial time interval allotted for transmission, provided that the bit budget for that frame’s GOF is less than the total size of the base layer NAL units in the GOF and provided that the bit budget allocated to that particular frame is of sufficient size to accommodate both the base layer and enhancement layer NAL units for that frame. Demircin, [0128]-[0131], FIGS. 7-8. Demircin’s process later repeats for the base layer picture (BHP) like that indicated by the gold shading in the annotated version of FIG. 1 above (and its GOF), and the gold base layer picture (BHP) would be subsequently transmitted following the path through steps 20, 26, 28, 30 and 31 in FIGS. 7-8, for example. Demircin, [0128]-[0131], FIGS. 7-8.



Demircin, FIGS. 7 and 8 (annotated).



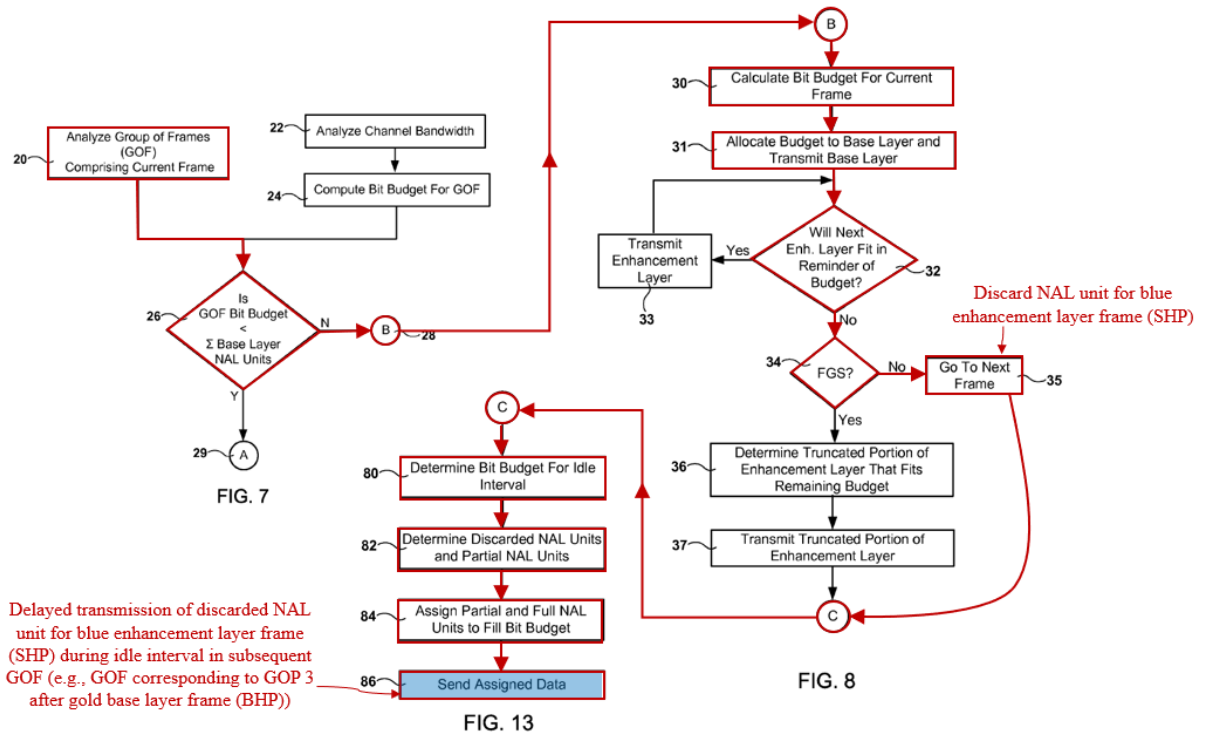
Demircin, FIGS. 7 and 8 (annotated).

191. *Demircin's Second Example Scenario: Delayed Transmission of Enhancement Layer NAL Unit(s)*

192. Considering the second example scenario in Demircin which an NAL unit for an enhancement layer picture (SHP) is transmitted, and hence received, on a delayed basis after the ordinary time of transmission/reception for the GOP to which the enhancement layer picture belongs, it is clear that the receiver would receive the blue-shaded SHP identified in the annotated version of FIG. 1 (above) after the gold-shaded BHP—particularly when transmission of the blue-shaded SHP is delayed by at least five frame transmission intervals.

193. Relevant to this scenario, Demircin describes techniques for “Delayed NAL Unit Transmission.” Demircin, [0104]; [0104]-[0122], [0123]-[0128], FIGS. 13-14. With these techniques, “when the system detects there may be additional channel bandwidth available, transmission of NAL units that were previously discarded or transmitted only partially may be considered and performed” and “[e]ligible NAL units may be reconsidered for transmission in delayed manner.” *Id.*, [0034] (“This method results in high video quality because it increases channel bandwidth utilization.”). In particular, the system “may detect idle channel intervals (instances when the channel is not fully utilized), and subsequently transmit previously dropped or truncated NAL units whose decoding deadlines have not yet expired.” *Id.*, [0105]; *see also id.*, [0106]-[0108], [0148]-[0149].

194. Thus, consistent with Demircin's description of "Delayed NAL Unit Transmission," the blue-shaded SHP identified in the annotated version of FIG. 1 (above) would be transmitted by the transmitter and received by the receiver after the gold-shaded BHP when the blue-shaded SHP is initially discarded due to insufficient bit budget, but is later transmitted/received in a delayed fashion during a suitable idle interval detected after transmission/receipt of the gold-shaded BHP (which is transmitted without delay). Such delay in transmission by the transmitter, and hence receipt by the receiver, of the blue-shaded SHP would occur, for example, following the path through steps 20, 26, 28, 30, 31, 32, 34, and 35 of Demircin's FIGS. 7-8 for the blue-shaded SHP—which results in the blue enhancement layer frame initially being discarded/not transmitted during its ordinary transmission interval—and then following the path through steps 80, 82, 84, and 86 in FIG. 13 for delayed transmission of the previously discarded blue-shaded SHP during a detected idle interval. Demircin, [0128]-[0134], [0148].



Demircin, FIGS. 7-8, 13 (annotated).

195. To summarize Demircin's teachings above, Demircin prioritizes the transmission of base layer NAL units over enhancement layer NAL units. First, at 26, if the GOF bit budget is greater than or equal to the base layer NAL units, then the transmitter transmits the base layer (31). Demircin, [0131]. If the next enhancement layer does not fit within the bit budget (32) and fine grain scalability is not supported (34), then the enhancement layer frame data are not transmitted yet (35). *Id.*, [0131]-[0132]. Later, during the idle interval, discarded units for the previously transmitted frames are sent (80-86). *E.g., id.*, [0149]. Therefore, in Demircin, enhancement layer frame data may be transmitted in a delayed manner relative to corresponding base layer

frames. And has already been established, Demircin transmits base layer and enhancement layer frames in units of GOP.

196. The amount of delay incurred by a discarded but later-transmitted enhancement layer NAL unit will depend on a number of factors such as how soon idle interval(s) occur after the initial transmission opportunity for the discarded unit, the bit budget available in the idle interval(s), and the priority of the discarded NAL unit relative to other discarded or truncated NAL units that are eligible for delayed transmission. However, it would have been obvious that Demircin's system would allow for the discarded blue-shaded SHP to be transmitted in a delayed fashion during an idle interval *after* transmission of the gold-shaded BHP. Demircin explains that discarded NAL units can be transmitted on a delayed basis so long as their "decoding deadlines have not yet expired." Demircin, [0105]. A "decoding deadline" refers to the time when a frame that has been received by a decoding device is scheduled to be retrieved from a coded frame buffer and decoded. SAMSUNG-1034, 9-10, 15-16. If both the base and enhancement layer NAL units for a frame are available in the coded frame buffer by the decoding deadline for that frame, then the decoder decodes the enhancement layer picture with reference to the base layer picture (and potentially other pictures). But if only the base layer NAL unit for a frame is available in the buffer by the frame's decoding deadline, then the decoder can only use the base layer NAL unit to decode that frame at a lower quality, for example.

197. Decoding deadlines, as well as “playback” or “display” deadlines, in streaming video systems depend in large part on how long the receiver buffers coded frame data before the frames are decoded and displayed. Long buffering times typically lead to performance improvements and reduce the risk of coded frame data being unavailable by the time that a frame is to be decoded and displayed, albeit at the expense of delayed playback and larger buffer size requirements. SAMSUNG-1034, 9-10, 15-16. Notably, buffering durations of 5-15 seconds were common before the Critical Date. *Id.* With so much time available before a frame is decoded, Demircin’s transmitter would have had ample ability to delay transmission of the NAL unit for a discarded enhancement layer picture (e.g., the blue SHP) past transmission of the NAL unit for a base layer picture (e.g., the gold BHP) whose coding order and display order are later by a factor of a group of pictures than the coding order and display order of the enhancement layer picture (e.g., the blue SHP) before the decoding deadline for the discarded enhancement layer picture (e.g., the blue SHP) expires. Thus, even when a frame’s enhancement layer picture is transmitted well after that frame’s base layer picture (e.g., delayed by ~5-15 seconds), the receiver of a streaming video program having a typical frame buffer can still receive the enhancement layer picture in sufficient time to decode it with reference to the earlier-received and buffered base layer picture before the decoding deadline expires.

198. Demircin does not otherwise restrict use of its techniques for delayed NAL unit transmissions, despite noting that other embodiments where “NAL units may have only one transmission opportunity at their first scheduled transmission ... perform well when the initial buffering duration is relatively long, for example longer than 300 ms[]” Demircin, [0104]. This disclosure does not preclude use of delayed NAL transmissions when the initial buffering duration is relatively long, and in any event, a POSITA would have been motivated to use delayed NAL unit transmissions in many cases when the initial buffering duration is relatively long to provide additional opportunities to transmit enhancement layer pictures that could not be initially transmitted if decoding deadline has not expired (such that the receiver could receive and decode the enhancement layer picture to render higher quality video).

199. Still, Dermicin meets and renders obvious the requirements of claim 1 in the scenario where the buffering duration is shorter (e.g., less than 300 ms). In this case, if the enhancement layer NAL units for a frame are received in the buffer within 300 ms of arrival of the base layer picture for the frame, then the enhancement layer picture can be decoded for that frame. It follows that, for transmission of frames in display order, the blue-shaded SHP would need to be delayed by five frame transmission intervals to be transmitted, and hence received, after the gold-shaded BHP. At a frame rate for decoding and presentation of 30 frames-per-second (FPS), this would implicate a delay of just ~167 ms—i.e., far below the 300 ms buffering

duration and well before the decoding deadline for frame 0 in display and decoding order. *Id.* At a frame rate for decoding and presentation of 60 FPS, this would implicate a delay of just ~ 83.3 ms—i.e., again, far below the 300 ms buffering duration and well before the decoding deadline for frame 0. Certainly, if a common buffering duration such as 5-15 seconds applied, the delays of ~ 167 or ~ 83.3 ms would be far less than the buffering duration and the decoding deadlines would not have expired.

200. Similarly, to the extent frames are transmitted in decoding order, the blue-shaded SHP would need to be delayed by seven frame transmission intervals to be transmitted, and hence received, after the gold-shaded BHP. SAMSUNG-1003, ¶198. At 30 frames-per-second (FPS), this would require a delay of ~ 233.3 ms—i.e., still well below the 300 ms buffering duration and decoding deadline for frame 0. *Id.* At 60 FPS, this would require a delay of just ~ 116.7 ms—i.e., again, far below the 300 ms buffering duration and well before the decoding deadline for frame 0 in display and decoding order.

201. For each of these reasons, Demircin discloses and renders obvious a receiver that receives an enhancement layer picture such as the blue-shaded SHP indicated above (supplementary hierarchical picture) whose coding order and display order are earlier by a factor of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures than those of a base layer picture such as the gold-shaded BHP indicated above (basic hierarchical picture) included in the base

layer data (basic video image coding data). Demircin thus renders this limitation obvious regardless of whether the supplementary hierarchical picture can be received before or must be received after the supplementary hierarchical picture. *Id.*

[1.5.3] a basic hierarchy and a supplementary hierarchy being set in units of the group of pictures;

202. Demircin discloses and renders obvious a basic hierarchy and a supplementary hierarchy being set in units of groups of pictures. For example, in FIG. 1, Demircin illustrates a “GOP structure and layering structure of a sample SVC bit stream.” Demircin, [0036]. Note that MPEG states that a GOP is “the smallest coding unit that can be independently decoded within a sequence.” SAMSUNG-1015, 54. As discussed above and shown below, in FIG. 1, the “GOP size is equal to 4 frames” and the depicted structure “uses I-frames 2, which are intra-predicted from within the frame; B-frames 3, which may be predicted from one or more other frames,” and optionally “P-frames 4 which are predicted from another I-frame.” Demircin, [0036]; *see also id.*, [0039], [0043], [0095].

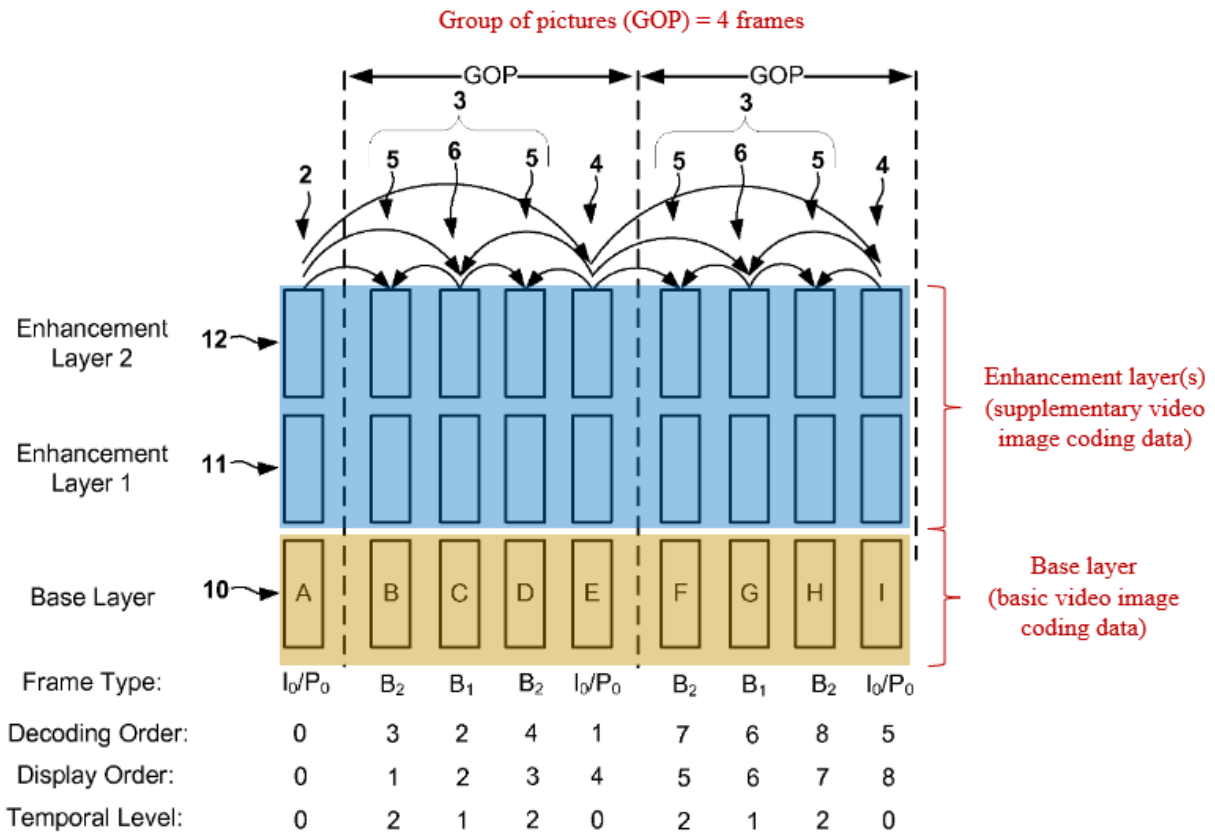


FIG. 1

Demircin, FIG. 1 (annotated).

203. The frame structure depicted in FIG. 1 is based on Scalable Video Coding (SVC), which provides a basic hierarchy and a supplementary hierarchy. Demircin, [0035]-[0036]. Demircin teaches that “a base layer provides a minimum level of video quality” and “[o]ne or more enhancement layers provide additional quality in an incremental manner, also referred to as progressive refinement.” *Id.*, [0035]. “Decoding of the first enhancement layer requires availability of the base layer” and “[d]ecoding of each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” *Id.* “Furthermore, temporal scalability

is supported based on a flexible temporal prediction structure, for example using hierarchical B frames.” *Id.* “Temporal scalability utilizes the concept of a temporal level, which indicates the level of a video frame in a hierarchical temporal prediction structure.” *Id.* Demircin thus discloses and renders obvious basic and supplementary hierarchies between base layer 10 and enhancement layers 11 and 12, and basic and supplementary hierarchies between lower temporal level(s) and higher temporal level(s). In each case, the basic and supplementary hierarchies are set in units of the group of pictures (*e.g.*, four frames across each of layers 10, 11, 12 as shown in FIG. 1). *Id.* For instance, like the basic hierarchies in the ’303 Patent, Demircin’s base layer 10 implements a hierarchy of intra- and inter-prediction coded pictures that can be decoded in the presence of only base layer pictures, while Demircin’s enhancement layers 11, 12 each and collectively implement a hierarchy of intra- and inter-prediction coded pictures that are decoded with reference to lower/base layer pictures and other enhancement layer pictures. ’303 Patent, 9:57-67.

[1.6] acquiring basic video image coding data received before supplementary video image coding data that has been received at the moment; and

204. Demircin discloses and renders obvious the features recited in [1.6]. For context, the specification of the ’303 Patent provides disclosure relevant to this limitation in which basic video image coding data is “acquir[ed]” from memory:

Meanwhile, the coding data reconstruction unit 202 *stores the basic video image coding data provided from the receiving unit 200 in a*

memory inside the coding data reconstruction unit 202 (S1102). Alternatively, the coding data reconstruction unit 202 may store the basic video image coding data in an external memory. In short, the coding data reconstruction unit 202 is simply required to be able to *acquire basic video image coding data received before supplementary video image coding data that has been received at the moment.*

'303 Patent, 7:34-43 (emphasis added).

205. Demircin discloses and renders obvious similar actions of receiving base layer NAL units for one or more frames (*basic video image coding data*) at an earlier time, storing the base layer NAL units in memory (*e.g.*, a buffer), later receiving enhancement layer NAL units for the same one or more frames (*supplementary video image coding data*) at the moment, and acquiring the base layer NAL units for the one or more frames from the memory (*e.g.*, the buffer) for use in reconstructing video image coding data such that the frames can be decoded and rendered to yield higher-quality pictures than if the enhancement layer NAL units had not been received.

206. For example, Demircin describes “Delayed NAL Unit Transmission” techniques. Demircin, [0104]; [0104]-[0122], [0123]-[0128], FIGS. 13-14. With these techniques, “when the system detects there may be additional channel bandwidth available, transmission of NAL units that were previously discarded or transmitted only partially may be considered and performed” and “[e]ligible NAL units may be reconsidered for transmission in delayed manner.” *Id.*, [0034] (“This method results in

high video quality because it increases channel bandwidth utilization.”). In particular, the system “may detect idle channel intervals (instances when the channel is not fully utilized), and subsequently transmit previously dropped or truncated NAL units whose decoding deadlines have not yet expired.” *Id.*, [0105]; *see also id.*, [0106]-[0108], [0148]-[0149].

207. Demircin allocates bit budgets to groups of frames (GOFs) and individual frames within GOFs, and prioritizes transmission of base layer NAL units before enhancement layer NAL units. *See, e.g.*, Demircin, FIG. 8 (steps 30-33) & [0131]-[0132]; *id.*, FIG. 7 (step 26) & [0130], FIG.12. If the GOF bit budget is less than the total size of all base layer NAL units in a GOF, Demircin discards all enhancement layer NAL units in the GOF and they are eligible for delayed transmission in idle intervals, even as all or some of the base layer NAL units are transmitted earlier in scheduled intervals. *Id.*, [0144] (“When all base layers do not fit in the GOF bit budget, all enhancement layer units may be discarded or omitted from transmission at the current time 70.”), [0129]-[0130], FIGS. 7, 12, 13. If the GOF bit budget can accommodate all base layer NAL units in a GOF but the bit budget for a particular frame is only sufficient to accommodate the frame’s base layer NAL unit, then that frame’s enhancement layer NAL unit(s) are discarded and provided for delayed transmission in idle intervals. *Id.*, [0132]-[0134]. In each of these examples, Demircin teaches that base layer NAL units for one or more frames (*basic video image coding data*) are transmitted before enhancement

layer NAL units for the one or more frames (*supplementary video image coding data*). NAL units transmitted earlier in time would be received at the receiver earlier in time than NAL units transmitted later, and it would have been obvious for Demircin's explicitly disclosed delay in transmitting enhancement layer NAL units to result in a situation where the receiver receives the base layer NAL units for one or more frames before the corresponding enhancement layer NAL units for those frames.

208. Demircin further discloses that the receiver stores received NAL units for frames that are to be later decoded in a "playout buffer." Demircin, [0065] ("playout buffer at the client"), [0096] ("playout buffer underflows"), [0104]. Demircin916, incorporated by reference in Demircin, explains that "Buffering of audio/video data is necessary and commonplace in audio/video transmission systems" and "[t]he receiving (decoding) part of the system typically includes a buffer at the input of the decoder." Demircin077, [0252], [0255] & FIG. 39 (decoder buffer 528). Demircin, through Demircin077, further teaches that "the receiver buffers the incoming data units from the wireless network before passing the data units to the decoder." *Id.*, [0261]. The receiver thus acquires data unit(s) for a frame from the buffer when the frame is to be reconstructed (*e.g.*, including decoding operations), and it would have been obvious in such cases for Demircin's receiver to acquire from the buffer base layer NAL units for one or more frames (*basic video image coding data*) that were received before the delayed enhancement layer NAL units for the one or more frames (*supplementary video image coding*

data) received at the moment—including, for example, when the delayed enhancement layer NAL units are received just in time “at the moment” the one or more frames are to be decoded. *See supra*, [1.5.2].

[1.7] reconstructing video image coding data from the basic video image coding data and the supplementary video image coding data.

209. Demircin discloses and renders obvious [1.7]. As described above with respect to [1.6], it would have been obvious for Demircin’s receiver to acquire from the buffer base layer NAL unit(s) for one or more frames (*basic video image coding data*) that were received before the delayed enhancement layer NAL units for those frames (*supplementary video image coding data*) received at the moment. *Supra*, [1.6]. Demircin also teaches Scalable Video Coding (SVC). Demircin, [0026], [0035]-[0036]. In SVC, “[a] base layer provides a minimum level of video quality, and may provide compatibility with H.264/AVC.” *Id.*, [0035]. “One or more enhancement layers provide additional quality in an incremental manner, also referred to as progressive refinement.” *Id.* “Decoding of the first enhancement layer requires availability of the base layer” and “[d]ecoding of each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” *Id.* Thus, Demircin discloses and/or it would have been obvious that, when the receiver receives the base layer and enhancement layer NAL units for a frame before the expiration of the decoding deadline, the receiver reconstructs higher-quality video image coding data by combining the base layer and enhancement layer NAL units consistent with SVC techniques. *Id.*

For example, if the scalability were temporal, Demircin's scheme would transmit the I-frames and P-frames within a bit budget as high-priority data, (and could reconstruct video slowly from those frames), then if idle time permitted transmission of interstitial B-frames, Demircin's scheme would reproduce those frames using motion-compensated prediction from the previously received I- and P-frames. Similar examples apply to SNR and spatial scalability in which the enhancement layers are predicted from the base layer. Recall that enhancement layer frames require corresponding base layer frames for decoding; therefore, to the extent that enhancement layer frames are transmitted, their decoding and reconstruction would rely on corresponding base layer frames. Therefore, Demircin teaches reconstructing video image coding data from the basic video image coding data and the supplementary video image coding data.

2. Claim 2

[2.P]-[2.2] A video image coding data transmission and reception system comprising: a video image coding data transmitter that codes and transmits a video image; and a video image coding data receiver that receives and reconstructs video image coding data,

210. *Supra*, [1.P].

[2.3.1] wherein the video image coding data transmitter includes: a first processor; and

211. Demircin's source/server (*video image data transmitter*) would have included a processor for substantially the same reasons as Demircin's receiver's processor would

have been obvious as discussed above in [1.1]. Demircin, Abstract, [0019], [0025]-[0026]; SAMSUNG-1006, [0061], [0063].

[2.3.2] a first memory unit having instructions stored which, when executed by the first processor, cause the first processor to perform operations comprising:

212. Demircin's source/server (*video image data transmitter*) would have include a first memory unit having instructions executable by a first processor for substantially the same reasons as it would have been obvious for Demircin's receiver's to include a memory unit having executable instructions as discussed above in [1.2]. Demircin, Abstract, [0019], [0025]-[0026]; SAMSUNG-1006, [0061], [0063].

[2.4] acquiring the transmission rate of a network;

213. Demircin discloses "methods and systems for adapting a video transmission rate." Demircin, Abstract. Demircin discloses that the source/server "determine[s] the bandwidth of the communication medium over which a signal may be sent" and "[b]ased on the available bandwidth, a GOF bit budget may be computed 24." Demircin, [0129]. "The bit budget may represent an estimate of the amount of data that may be transmitted across the communication medium at the time of the estimate." *Id.*

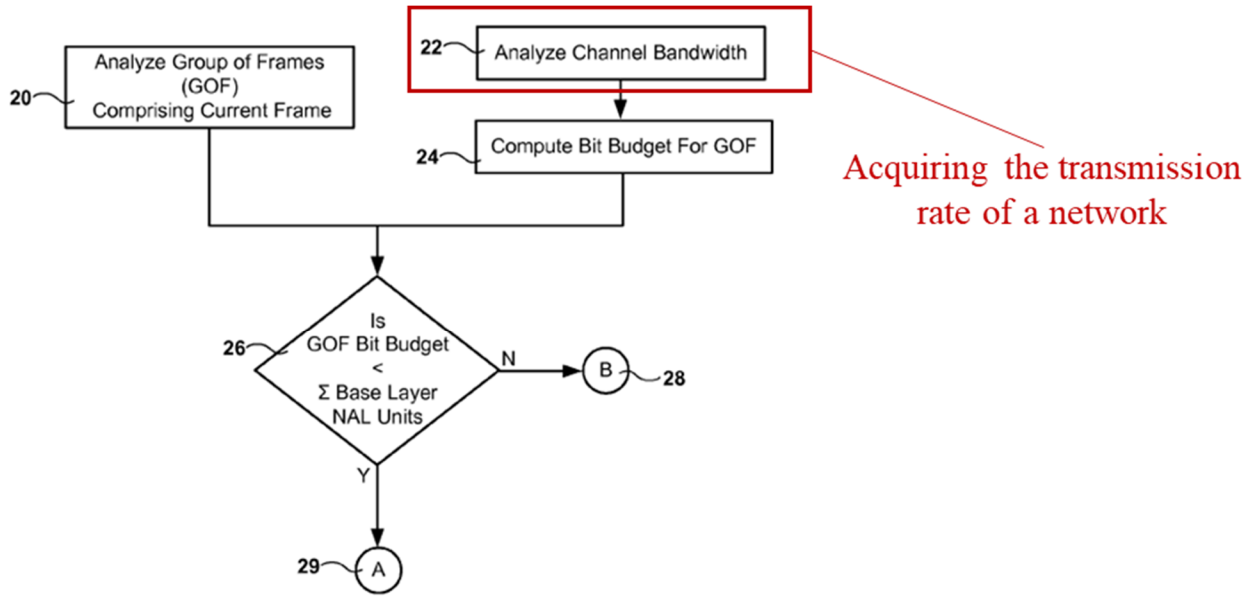


FIG. 7

Demircin, FIG. 7 (annotated).

[2.5] setting a transmission structure including a basic hierarchy and a supplementary hierarchy, the basic hierarchy and the supplementary hierarchy being set in units of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures;

214. *Supra*, [1.5.2]-[1.5.3].

[2.6.1] transmitting basic video image coding data of the basic hierarchy,

215. *Supra*, [1.3].

[2.6.2] wherein the first memory unit is further configured to store supplementary video image coding data of the supplementary hierarchy;

216. The source/server of Demircin would have had a memory for storing the NAL units for enhancement layer pictures (*supplementary video image coding data*), particularly in view of Demircin teaching that enhancement layer pictures can be transmitted on a delayed basis in idle intervals. Demircin, [0104]-[0106], FIGS. 12-13.

It would have been obvious that the discarded pictures would be stored so as to be available for delayed transmission, and the storage of the discarded pictures, like much other digital content, would have been in a memory unit.

[2.7] transmitting the supplementary video image coding data stored in the memory unit; and

217. *Supra*, [1.5.1]-[1.5.3].

[2.8] controlling the transmission of the supplementary video image coding data according to the transmission rate, and

218. Demircin's system determines whether to transmit enhancement layer data for frames in a GOF, and thus controls the transmission of the enhancement layer data for frames in a GOF, based on a bit budget for the GOF calculated based on the acquired bandwidth (i.e., transmission rate) of the network. Demircin, [0129], FIGS. 7-9; *supra*, [2.4].

219. Demircin further describes techniques for "Delayed NAL Unit Transmission," which involves controlled transmission of supplementary video image coding data. *Supra*, [1.P], [1.5.1]-[1.5.3]; Demircin, [0104]-[0122], [0123]-[0128], FIGS. 7-8, 12-14. With these techniques, "when the system detects there may be additional channel bandwidth available, transmission of NAL units that were previously discarded or transmitted only partially may be considered and performed" and "[e]ligible NAL units may be reconsidered for transmission in delayed manner." *Id.*, [0034].

[2.9] wherein the video image coding data receiver includes:

220. *Supra*, [1.P].

[2.10] a second processor; and

221. *Supra*, [1.1].

[2.11] a second memory unit having instructions stored which, when executed by the second processor, cause the second processor to perform operations comprising:

222. *Supra*, [1.2].

[2.12] receiving the basic video image coding data;

223. *Supra*, [1.3].

[2.13] decoding the received basic video image coding data so as to reproduce a video image;

224. *Supra*, [1.4].

[2.14] receiving the supplementary video image coding data including a supplementary hierarchical picture having coding order and display order that are earlier by a factor of the group of pictures than those of a basic hierarchical picture included in the basic video image coding data;

225. *Supra*, [1.5.1]-[1.5.3].

[2.15] acquiring basic video image coding data received before supplementary video image coding data that is currently received; and

226. *Supra*, [1.6].

[2.16] reconstructing video image coding data from the basic video image coding data and the supplementary video image coding data.

227. *Supra*, [1.7].

IX. Ground 2 – Claim 1 Is Obvious Under 35 U.S.C. § 103 based on Kimoto in Combination with Demircin

228. Kimoto in combination with Demircin renders claim 1 obvious and *invalid*.

A. Combination of Kimoto and Demircin

229. Kimoto describes techniques for coding, transmitting, and reproducing hierarchical video coding data including lower layer data and upper layer data. Kimoto, [0037]-[0045]. More specifically, Kimoto teaches receiving, storing, and even displaying the hierarchically structured lower layer data earlier in time; then, later in time, receiving hierarchically structured upper layer data, and synthesizing the upper layer data with corresponding stored lower layer data to generate high quality data. Kimoto, [0069]-[0070]. Put simply, Kimoto teaches substantially the same features as claim 1 of the '303 Patent. The only notable difference here is that Kimoto does not expressly disclose setting the hierarchically coded lower and upper data as groups of pictures (GOPs). However, Kimoto implicitly refers to GOPs when referring to the application of the invention to the MPEG scalable standards: "The international standards of the hierarchical coding methods include MPEG-2 scalable, MPEG-4 scalable, MPEG-4 FGS, H.264/MPEG-4 AVC scalable video coding, and Motion JPEG 2000. The present invention is applicable to such hierarchical coding methods illustrated and other hierarchical coding methods.". Kimoto [0038]. Therefore it would have been obvious to combine Kimoto's system with that of Demircin, as described in detail below. In

the combination, the hierarchically coded lower and upper data of Kimoto are set in GOPs based on Demircin. *Id.* For example, a GOP like that depicted in Demircin's FIG. 1 (which contains an intra-coded picture and multiple inter-prediction coded pictures) would have been obvious to apply to Kimoto's system in the combination, in an example. *Id.* A POSITA would have found it obvious to modify the lower and upper layer data of Kimoto to be set in units of GOP, in a manner similar to Demircin, to achieve the predictable results of accurate and efficient synthesis of newly received upper layer data with stored lower layer data, and to support the ability of delayed playback described in Kimoto.

B. Reasons to Combine Kimoto and Demircin

230. There are several reasons a POSITA would have been motivated to modified Kimoto at the time so that Kimoto's lower and upper layer data would be organized in GOPs consisting of an intra coded picture (*e.g.*, an I-frame) and a plurality of inter-prediction coded pictures (*e.g.*, hierarchical B-frames), consistent with the structure depicted in FIG. 1 of Demircin, for example. At the outset, Kimoto makes it clear that the invention is applicable to hierarchical coding methods including "MPEG-2 scalable, MPEG-4 scalable, MPEG-4 FGS, H.264/MPEG-4 AVC scalable video coding, and Motion JPEG 2000." Kimoto, [0038]. As discussed above, setting hierarchical pictures into GOPs was well known and used as early as 1993. *E.g., supra*, §§ II.D, V.5. A POSITA would have found it obvious to set the frames of Kimoto into

units of GOP because doing so would have been standard practice or commonplace when following the standard practice, and for other reasons described in this declaration.

231. Moreover, Kimoto already describes that its hierarchical structure includes a non-reference frame and a plurality of corresponding reference frames. *E.g.*, Kimoto [0056]. Thus, the underlying frame structure for intra- and inter-prediction in Kimoto already includes the same number of frames necessary to satisfy the claim limitation of the '303 Patent. *See* '303 Patent, 9:33-36 (describing reference frames). Per H.264 2007, a non-reference picture is not used for inter prediction of any other pictures. SAMSUNG-1009, 9. A reference picture is used for inter prediction in the decoding process of subsequent pictures in the decoding order. *Id.*, 11. Also note a picture is a collective term for field or frame. *Id.*, 10. MPEG defines picture as source, coded or reconstructed image data. SAMSUNG-1015, 8. Kimoto's affirmative applicability to the H.264 and MPEG standards, coupled with the described frame structure, would have suggested to a POSITA to set the frames in GOPs for organization, transmission, decoding, and synthesis. Moreover, recall that Kimoto is about reconstructing a video at a later time, after a first viewing of lower layer data. Thus, using GOP would have also aided in the ability of a viewer to perform random access viewing of the synthesized data. GOPs were designed for random access, and a POSITA trying to reconstruct a

video for subsequent (and non-real-time) viewing would have recognized the advantages of GOP for implementing random access in video playback.

232. The MPEG standards also describe GOPs in terms of intra-coded pictures (I-frames) and inter-prediction coded pictures (B-frames). *E.g.*, SAMSUNG-1015, 15, 54, 68. A GOP includes a number of frames starting with an intra-coded picture (also known as key frames) until the next intra-coded picture. *Id.* The patent owner's initial infringement contentions identify a GOP as "a grouping of a series of frames into sequences." *E.g.*, SAMSUNG-1017, 16.

233. According to MPEG1, a GOP may be of any length. The length of the GOP is flexible and may be tailored to an application. GOP length can be selected based on the compression rate, the desired video quality, the bandwidth for transmission, the decoding time and resources, etc. The shorter the GOP, the fewer inter-prediction coded pictures, which usually means higher bitrate. The longer the GOP, the more inter-prediction coded pictures, which usually means higher compression. Therefore, the size and structure of the GOP is adaptable based on various implementation scenarios. Applications requiring random access, fast-forward playback, or fast and normal reverse playback may use relatively short groups of pictures. Groups of pictures may also be started at scene cuts or other cases where motion compensation is ineffective. SAMSUNG-1015, 16.

234. As I have shown in the general discussion of the prior art, it was well known prior to the Critical Date that a GOP structure included one intra-coded picture (one I-frame) and a plurality of inter-prediction coded pictures (B-frames). *E.g., supra* §§ II.D.1-4, V.A.5, & VI.A. Specifically, Demircin teaches a GOP that includes an intra-coded picture (I-frame) and three inter-prediction coded pictures (B-frames). *E.g., Demircin, FIG. 1.*

235. Next, organizing pictures in the lower and upper layer data into GOPs would aid Kimoto in decoding a video stream by allowing it to recognize sets of pictures that can be decoded without reference to any other pictures in the stream. Setting lower and upper layer data into GOPs can improve the desired synthesis of lower and upper layer data in Kimoto to quickly begin reconstructing high-quality data based on various known locations of keyframes (i.e., I-frames) in the video stream. In addition, organizing pictures in the lower and upper layer data into GOPs would increase the efficiency and reliability of the picture alignment between the upper layer data and the corresponding lower layer data during reproduction.

236. Organizing pictures in the lower and upper layer data into GOPs that each consist of a series of inter-prediction coded pictures (*e.g., hierarchical B-frames*) together with an intra-coded picture (*e.g., I-frame*) like that depicted in Demircin's FIG. 1, would beneficially reduce the size of the video stream, particularly since the inter-

prediction coded pictures typically undergo greater compression and include less data than the intra-coded picture.

237. Organizing pictures in the lower and upper layer data into GOPs of size four like that shown in FIG. 1 of Demircin would have optimized one or more characteristics of the video stream (although smaller or larger GOP sizes would also be suitable in the combination. The selection of GOP size and IBP (frame) pattern would have been obvious design choice for Kimoto to achieve lower and upper layer data transmission, particularly in view of Kimoto's desire to compensate for reproduction delays caused, in part, by transmission bit rate limitations. Shorter GOPs can preserve more detail in high-motion data, for example, while longer GOPs improve the amount of compression that can be achieved since intra-coded frames occur less frequently. For example, Kimoto relies on delay times and transmission bandwidths to determine the length of the upper layer first data and upper layer subsequent for transmission. *E.g.*, Kimoto, claim 42. Thus, Kimoto would have sized the GOP with transmission bandwidth in mind.

238. A POSITA would have been motivated to turn to Demircin specifically because both Kimoto and Demircin concern prioritizing the transmission of base layer (or lower layer) data over the enhancement layer (or upper layer) data, and the synthesis of lower and upper layer data later in time. Both Kimoto and Demircin involve prioritizing base or lower layer frames for transmission and storage before enhancement

or upper layer frames. *E.g.*, Kimoto, [0041]-[0042], [0067]; Demircin, [0131]-[0132]. And both Kimoto and Demircin describe synthesizing high-quality data at the receiver by reconstructing lower layer data with upper layer data. *E.g.*, Kimoto, [0069]; Demircin, [0035]. Demircin provides structural information in the form of GOPs that Kimoto can use to accomplish the same goal of prioritizing the transmission of lower layer data, while also transmitting upper layer data, and synthesizing lower and upper layer data to reconstruct a high-quality video. So using Demircin would have benefited Kimoto to achieve Kimoto's desired goals.

239. Kimoto is silent as to the details of how that synthesis works, and a POSITA would have been motivated to seek a solution, like that taught by Demircin, which includes setting the lower and upper layer data into GOPs. Using GOPs allows the stored lower layer data to be aligned quickly and efficiently with corresponding upper layer data for synthesis and reproduction. In addition, using GOPs aids Kimoto in decoding a video stream by allowing it to recognize sets of pictures that can be decoded without reference to any other pictures in the stream. Therefore, a POSITA seeking to implement the prioritization scheme of Kimoto would have been motivated to seek out Demircin and to modify Kimoto using the GOP teachings of Demircin. The GOP of Demircin already included one intra-coded picture and a plurality of inter-prediction coded pictures, which was a commonplace GOP structure at the time as discussed at length above. *E.g.*, Demircin, FIG. 1, [0036], *supra*, §§ II.D, V.A.5, & VI.A.

Thus, the combination of Kimoto with Demircin's GOP renders the '303 Patent obvious and invalid.

240. Setting the lower and upper layer data of Kimoto in GOPs like that of Demircin, would have been met with a reasonable likelihood of success considering that the use of GOPs was a well-known video coding technique at the time. *See, e.g., supra* §§ II.D, V.A.5, & VI.A. First, Kimoto already explains that its “content data” is “hierarchically coded” and is “scalable in one or a plurality of directions including time (frame rate), space (resolution), and image quality (bit rate).” *Id.*, [0037]. Kimoto is broadly applicable to various standards, including MPEG-2 scalable, MPEG-4 scalable, MPEG-4 FGS, H.264 / MPEG-4 AVC scalable video coding, and Motion JPEG 2000. *Id.*, [0038]. Likewise, Demircin involves hierarchical scalable video coding. *E.g.,* Demircin [0035] (citing H.264/MPEG AVC/SVC).

241. Also, Kimoto describes reference frames and non-reference frames, explaining that certain frames in the coded data are “referred to when decoding other frames” while other frames are not referred to for purposes decoding other frames.⁶ *Id.*, [0056]-[0059], [0063]. While Kimoto is broadly applicable to various standards, including MPEG-2 scalable, MPEG-4 scalable, MPEG-4 FGS, H.264 / MPEG-4 AVC scalable video coding, and Motion JPEG 2000, to the extent that Kimoto does not

⁶ According to H.264 standard, “A reference frame may be used for inter prediction when P, SP, and B slices of a coded frame are decoded. See also reference picture.” SAMSUNG-1009, 32.

expressly describe how pictures within the lower and upper layers could be formally grouped for prediction purposes, grouping the lower and upper layer pictures would have been a desirable design choice because doing so was common practice in the field of hierarchical and scalable video coding, and doing so would have benefited Kimoto's transmission and reconstruction of upper and lower layer data.

242. Grouping pictures in hierarchical and scalable video streams was well known before the Critical Date as demonstrated by Demircin and other references. As described above (*supra*, §§ VI.A & VIII.A.1), Demircin discloses “an exemplary prediction structure, GOP structure and layering structure of a sample SVC bit stream” that includes a base layer and one or more enhancement layers. Demircin, [0036], FIG. 1. Demircin's structure as shown in FIG. 1 groups frames in the SVC stream into groups of pictures (GOPs) each having four frames, including an intra-coded picture (I-frame) and a set of three bi-directional inter-prediction coded pictures (hierarchical B-frames). *Id.*, [0036]-[0037], FIG. 1. It would have been obvious to combine the teachings of Kimoto and Demircin, including modifying Kimoto's system in view of Demircin so that Kimoto's lower and upper layer data would be organized in GOPs consisting of an I-frame and a plurality of inter-prediction frames (*e.g.*, hierarchical B-frames), *e.g.*, consistent with the structure depicted in FIG. 1 of Demircin.

243. The combination of Kimoto and Demircin invalidates claim 1 of the '303 Patent because the combination teaches all of the elements of the claim. The use of a

GOP in Kimoto would have been met with a reasonable expectation of success because the use of GOPs in video image coding was a well-known, standard practice at the time. The application of known techniques for organizing a video stream into GOPs, as taught in Demircin, would have improved the similar type of scalable video coding system of Kimoto in a similar way, which would have yielded no more than predictable results. Further, combining the teachings of Kimoto and Demircin in the manner described above would have been met with reasonable expectation of success. Grouping frames into GOPs had been well-established in hierarchical and scalable video coding before the Critical Date. *E.g., supra* §§ II.D, V.A.5, & VI.A. Implementing GOPs and frame structures like those disclosed in Demircin in other scalable video contexts such as that described in Kimoto would have been well within the norms of standard practice at the time. *Id.*; Kimoto, [0038]; Demircin, [0020], [0026].

C. Analysis

1. Claim 1

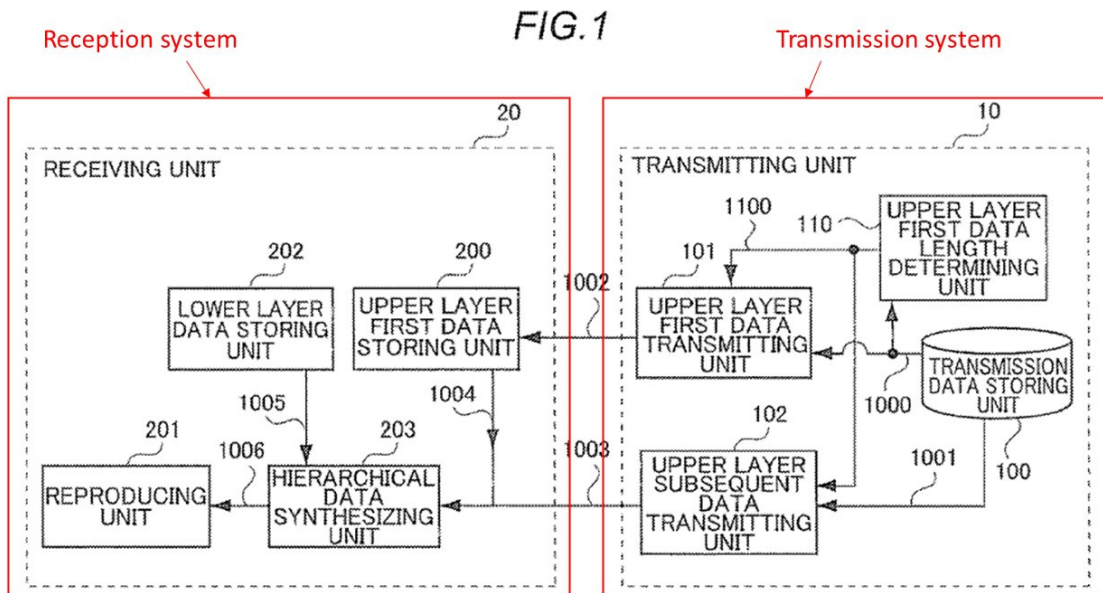
244. The Kimoto-Demircin combination renders claim 1 obvious and invalid.

[1.P] A video image coding data receiver comprising:

245. Assuming the preamble is deemed limiting, the Kimoto-Demircin combination renders it obvious. For example, in a first embodiment, Kimoto FIG. 1 illustrates a “moving image distribution system” that “includes a transmitting unit 10 and a receiving unit 20.” Kimoto [0035]. Kimoto teaches a receiving unit 20 (*video image*

coding data receiver) that receives lower layer and upper layer video image coding data from a transmitting unit 10 (*video image coding data transmitter*). Kimoto, [0043], [0047]. Kimoto

FIG. 1 illustrates a receiving unit that can perform image decoding:



Kimoto, FIG. 1 (annotated).

246. In more detail, Kimoto explains that “FIG. 1 is a block diagram showing a first embodiment of the moving image distribution system according to the present invention.” Kimoto, [0035]. “The transmitting unit 10 includes a transmission data storing unit 100, an upper layer first data transmitting unit 101, an upper layer subsequent data transmitting unit 102, and an upper layer first data length determining unit 110.” Kimoto [0036]. The transmitting unit codes video image coding data according to “international standards of the hierarchical coding methods include

MPEG-2 scalable, MPEG-4 scalable, MPEG-4 FGS, H.264/MPEG-4 AVC scalable video coding, and Motion.JPEG 2000. The present invention is applicable to such hierarchical coding methods illustrated and other hierarchical coding methods.” Kimoto [0038]. “The transmission data storing unit 100 contains data on reproduction contents. The content data stored is hierarchically coded.” Kimoto [0037]. The transmitting unit coded a video image that would be decoded at the receiver. Kimoto [0037]-[0038].

247. In addition, the transmitting unit 10 transmits first upper layer and subsequent upper layer data to the receiving unit 20. “The transmitting unit 10 includes a transmission data storing unit 100, an upper layer first data transmitting unit 101, an upper layer subsequent data transmitting unit 102, and an upper layer first data length determining unit 110.” Kimoto [0036]. “The upper layer first data transmitting unit 101 reads upper layer data as long as the upper layer first data length 1100 from the transmission data storing unit 100, and transmits the upper layer data to the receiving unit 20 (S.11).” Kimoto [0047].

248. The receiving unit 20 includes “an upper layer first data storing unit 200, a reproducing unit 201, a lower layer data storing unit 202, and a hierarchical data synthesizing unit 203.” *Id.*, [0039]. The receiving unit 20 is “capable of reproducing contents.” *Id.*, [0040].

249. To reproduce moving image contents, “the hierarchical data synthesizing unit 203 synthesizes the upper layer first data 1004 stored in above S03 and the corresponding part 1005 of the lower layer data to generate high quality data 1006 (S04).” *Id.*, [0069]. “The reproducing unit 201 then decodes the high-quality data 1006” to reproduce the content at the receiver. *Id.*, [0070] (emphasis added). Therefore, Kimoto’s receiving unit 20 receives video image coding data in the form of lower layer data and upper layer data from the transmitting unit 10, and reconstructs (i.e., synthesizes) video image coding data. (Note according to the ’303 Patent, “the basic video image coding data and the supplementary video image coding data of the coding data reconstruction unit 202” are synthesized for reproduction. *See* ’303 Patent, 12:51-56.)

[1.1] a processor; and

250. The Kimoto-Demircin combination renders element [1.1] obvious in view of Kimoto’s teaching of a processor. For example, according to Kimoto, “the moving image data distribution system may be implemented by hardware, software, or a combination of these.” Kimoto, [0111]. A system that was implemented in hardware and that executed software included a processor. FIG. 2 of Kimoto shows the “procedure of processing of the transmitting unit 10.” Kimoto [0044]. Structurally, the transmitting unit includes an upper layer first data length determining unit 110, and the

upper layer first data length determining unit 110 would have been implemented using a processor.

251. Kimoto also teaches that “a computer” can be made to “function as a receiving apparatus of a moving image data distribution system” Kimoto, [0024]. Computers included processors, and thus Kimoto’s receiving unit would also have included a processor. Indeed, Kimoto teaches that the receiving unit in each embodiment “performs the same *processing*.” Kimoto, [0108] (emphasis added). The receiving unit performed “processing” with a processor.

252. Kimoto would have implemented its receiving to include a processor in the receiving unit to carry out the various functions of the receiving unit as described in Kimoto and provided in the Kimoto-Demircin combination, including receiving, storing, retrieving, decoding, synthesizing, and reproducing moving images, as was commonly practiced at the time. Using a process was widely known at the time to implement image processing, storing, and reproduction, so using a processor in Kimoto would have been met by a reasonable expectation of success, especially in view of the common use of processors for carrying out processing functions in receiving units and other types of computers as of the Critical Date of the ’303 Patent.

[1.2] a memory unit having instructions stored which, when executed by the processor, cause the processor to perform operations comprising:

253. The Kimoto-Demircin combination renders element [1.2] obvious. For example, according to Kimoto, “the moving image data distribution system may be

implemented by hardware, software, or a combination of these.” Kimoto, [0111]. A system that was implemented in hardware and that executed software would have included a memory to store the software instructions.

254. Kimoto also teaches that a “computer” can be made to “function as a receiving apparatus of a moving image data distribution system” Kimoto, [0024]. Computers commonly included one or more memory units to store software instructions and other information, and thus Kimoto’s receiving unit would have included such a memory. Likewise, the transmitting unit 10 would have included a memory unit for storing instructions, including for calculating a “reproduction delay time for the case of transmitting only the upper layer data to the receiving unit 20, and determin[ing] an upper layer first data length 1100 (S10),” among other things. Kimoto [0047].

255. Kimoto additionally discusses caching of images, which would have used a cache memory for storing image data. *See, e.g.*, Kimoto, [0014]. Using a cache memory to store instructions for carrying out the various functions of the receiving unit as described in Kimoto and provided in the Kimoto-Demircin combination was a well know technique prior to the Critical Date.

256. To the extent Kimoto does not expressly disclose the recited memory unit, Kimoto would have included a memory in the receiving unit to store instructions for executing the various functions of the receiving unit as described in Kimoto and

provided in the Kimoto-Demircin combination. Including a memory in the receiving unit would have been met with a reasonable expectation of success implementing a processor in the receiving unit, especially in view of the common use of memory units for storing instructions in receiving units and other types of computers as of the Critical Date of the '303 Patent.

[1.3] receiving basic video image coding data;

257. The Kimoto-Demircin combination renders element [1.3] obvious. For example, Kimoto describes “lower layer data” (basic video image coding data) from which moving image contents can be viewed at a low quality. Kimoto [0025]. Kimoto explains that the “lower layer data on the reproduction contents is previously stored in the lower layer data storing unit 202 by the point in time when the contents are reproduced.” Kimoto [0041]. In situations where a user has already viewed low quality contents, “it is assumed that the low quality content data, or lower layer data, has already been transferred from a server to the receiving unit 20.” Kimoto [0042]-[0044], [0025]. Thus, in Kimoto, at some time before receiving the upper layer data, the receiving unit receives, stores, and even decodes and displays (a user views) the lower layer data. *Id.*

258. The lower layer data of Kimoto is hierarchically coded according to one of several international standards, including H.264/MPEG-4 advanced video coding (AVC) scalable video coding (SVC). *E.g.*, Kimoto, [0019]-[0024]; *e.g.*, Demircin, [0020]-[0021], [0026], [0035]-[0037]. According to both Kimoto and Demircin, hierarchical

coding of the video coding data (including lower layer data) can take different forms, including temporal scalability and signal-to-noise ratio (SNR) scalability (also referred to as image quality scalability). *E.g.*, Demircin, [0026]; Kimoto, [0037]. Kimoto teaches that the lower layer data and the upper layer data are hierarchically coded. *E.g.*, Kimoto, [0019]-[0024]. As discussed above, lower layer data reads onto basic video image coding data.

259. Based on the exemplary context illustrated by Demircin, in the combination, the hierarchical structuring of content data in Kimoto includes basic hierarchical pictures organized in units of GOPs. *Supra*, IX.A-B. As Demircin further explains, coarse-grained SNR scalability “is achieved using a layered structure for each frame, with a base layer 10 and two enhancement/refinement layers”—although “[m]ore or fewer enhancement layers may be used.” Demircin, [0037]. The following drawing (Illustration 1) based on Demircin FIG. 1 illustrates an example of SNR scalability for a known frame structure in the Kimoto-Demircin combination, which includes (i) a base layer containing lower layer or basic video image coding data and (ii) one or more enhancement layers containing upper layer or supplementary video coding data:

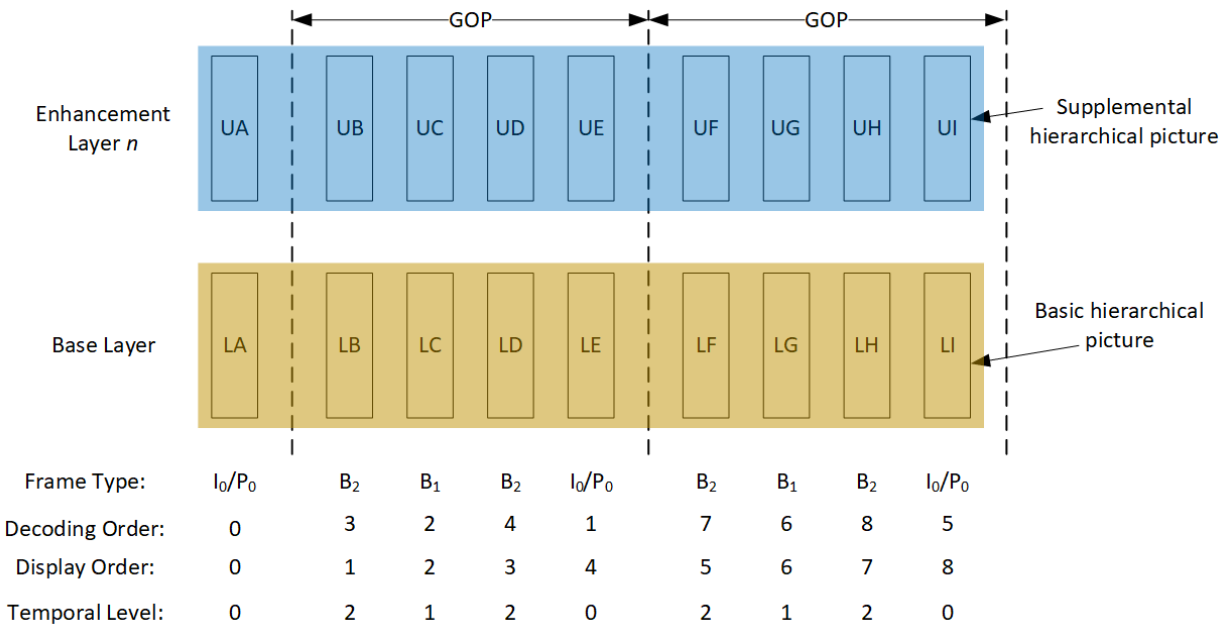


Illustration 1 of Kimoto-Demircin Combination

260. As shown in Illustration 1, the base (lower) layer includes an intra-prediction coded I-frame and a plurality of B-frames inter-prediction coded from one or more other frames. Demircin, [0036]; *see also id.*, FIG. 1. The I-frames and B-frames of the base layer are each basic hierarchical pictures. Demircin's disclosure of a base layer, which corresponds to the lower layer of Kimoto in the combination, is organized into groups of pictures (GOPs), where the "GOP size is equal to 4 frames." Demircin, [0036]. Illustration 1 further shows one enhancement layer n , though Demircin teaches that more than one enhancement layer can be used in coarse-grain SNR scalability. Demircin, [0037]. When performing decoding with SNR scalability, decoding the "first enhancement layer requires availability of the base layer," and decoding "each

subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” Demircin, [0035]..

[1.4] decoding the received basic video image coding data so as to reproduce a video image;

261. The Kimoto-Demircin combination renders element [1.4] obvious. For example, Kimoto discloses “a situation where a user *who has viewed low quality contents* on a mobile terminal such as a cellular phone and a PDA (Personal Data Assistant) views the same contents in high quality on the receiving unit 20 which is a part of the embodiment of the present invention.” Kimoto, [0042] (emphasis added). In such cases, “it is assumed that the low quality content data, or lower layer data, has already been transferred from a server to the receiving unit 20.” *Id.*, [0043]; *see also id.*, [0025] (“reduce the first data transmission time by transmitting only the upper layer data when low quality data has already been viewed on a receiving terminal that contains the lower layer data”).

262. In the combination, the lower layer data such as the base layer data for SNR scalability as taught in Demircin, and that Kimoto teaches “has already been transferred from the server to the receiving unit 20,” renders obvious the received basic video image coding data. Kimoto, [0043]. The receiving unit decodes the received lower layer data since coded video data is ordinarily, if not always, decoded for it to be viewed. *Id.*, [0041]-[0043].

263. Kimoto teaches that the lower layer data is stored in the lower layer data storing unit 202, *e.g.*, Kimoto, [0043], and when reproducing the higher quality contents, “the hierarchical data synthesizing unit 203 synthesizes the upper layer first data 1004 stored in above S03 and the corresponding part 1005 of the lower layer data to generate high quality data 1006 (S04)” stored in the lower layer storing unit. *Id.*, [0069]. Thus, in Kimoto, the lower layer data is decoded and viewed as low quality contents at one point in time; and later, after the upper layer data is received, the lower layer data is acquired from the lower layer data storing unit for reproduction as a higher quality content with the upper layer data. *Id.*

[1.5.1] receiving supplementary video image coding data including a supplementary hierarchical picture . . .

264. The Kimoto-Demircin combination renders element [1.5.1] obvious. To start, Kimoto teaches receiving upper layer data, which corresponds to “supplementary video image coding data” as claimed. Kimoto, [0067]; *supra* [1.3]. By way of explanation, in Kimoto, the upper layer data can be split into components to account for reproduction delays. *E.g.*, Kimoto, [0047]-[0049]. Kimoto refers to these components as upper layer first data and upper layer subsequent data. *E.g.*, Kimoto, [0047], [0068]. The upper layer data meets the claim limitation whether it includes upper layer first data, upper layer subsequent data, or both. In addition, Kimoto teaches that the upper layer data is hierarchically coded, meaning that the upper layer data is decoded

using the lower layer data using hierarchical coding techniques consistent with MPEG or H.264 standards. *E.g.*, Kimoto, [0019]-[0024], [0038].

265. Before the receiving unit 20 can synthesize the stored lower layer data with the upper layer data to reproduce high-quality contents, “the transmitting unit 10 transmits upper layer first data” to the receiving unit. Kimoto, [0045]. At the transmitter, the “upper layer first data length determining unit 110 calculates a reproduction delay time” for “transmitting only the upper layer data to the receiving unit 20.” Kimoto, [0047]. The upper layer first data length determining unit then “determines an upper layer first data length.” *Id.*

266. After the receiving unit receives, decodes, and reproduces the lower layer data (basic video image coding data), *e.g.*, Kimoto, [0042]-[0043], the “receiving unit 20 receives the upper layer first data 1002 from the transmitting unit 10, and stores the upper layer first data 1002 into the upper layer first data storing unit 200 (S02).” Kimoto, [0067]. The hierarchical data synthesizing unit “synthesizes the upper layer first data” with the “corresponding part” of the stored lower layer data to generate high quality data.” *Id.*, [0069]. Likewise, the hierarchical data synthesizing unit “subsequently synthesizes the upper layer subsequent data” with “the corresponding part” of the lower layer data to generate high quality data. *Id.*, [0070]. Since the upper layer first data and the upper layer subsequent data are each synthesized with the lower layer data, either or both read onto the claimed supplementary video image coding data. *Id.* Thus,

Kimoto discloses and the combination renders obvious receiving supplementary image coding data (i.e., the upper layer data).

267. Kimoto further discloses, and the combination renders obvious, that the upper layer data (whether upper layer first data, upper layer subsequent data, or both) includes a supplementary hierarchical picture. For example, Kimoto teaches non-reference pictures at the beginning of the contents of the upper layer data, *e.g.*, Kimoto, [0063], and reference pictures. *E.g.*, Kimoto, [0055]-[0057].

268. In the combination, the upper layer data is hierarchically coded using an international standard of hierarchical coding, such as H.264/MPEG-4 AVC SVC. *See, e.g.*, Kimoto, [0019]-[0024]; Demircin, [0020]-[0021], [0026], [0035]-[0037]. According to both Kimoto and Demircin, hierarchical coding of the video coding data (including upper layer data) can take different forms, including temporal scalability and signal-to-noise ratio (SNR) scalability (also referred to as image quality scalability). *E.g.*, Demircin, [0026]; Kimoto, [0037]. Kimoto teaches that the lower layer data and the upper layer data are hierarchically coded. *E.g.*, Kimoto, [0019]-[0024].

269. As Demircin further explains, coarse-grained SNR scalability “is achieved using a layered structure for each frame, with a base layer 10 and two enhancement/refinement layers”—although “[m]ore or fewer enhancement layers may be used.” Demircin, [0037]. The following drawing based on Demircin’s FIG. 1 illustrates an example of SNR scalability for a known frame structure in the Kimoto-

Demircin combination, which includes (i) a base layer containing lower layer or basic video image coding data and (ii) one or more enhancement layers containing upper layer or supplementary video coding data:

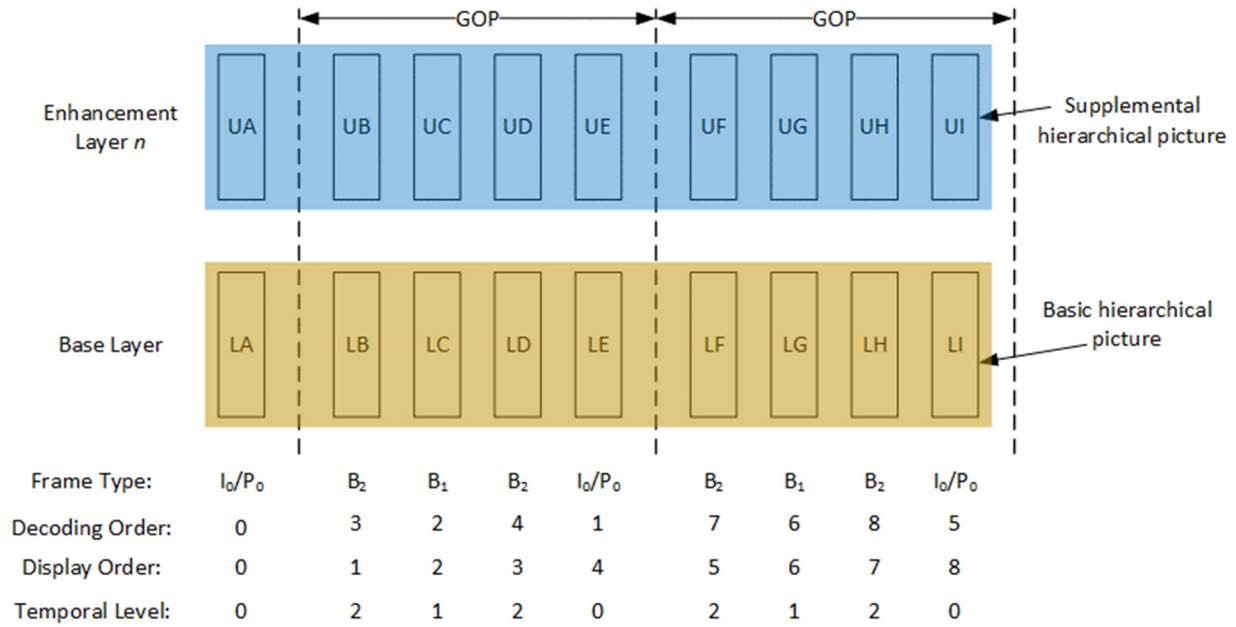


Illustration 1 of Kimoto-Demircin Combination.

Supra, [1.3].

270. The Illustration 1 illustration shows one enhancement layer n , though Demircin teaches that more than one enhancement layer can be used in coarse-grain SNR scalability. Demircin, [0037]. The orange highlighted boxes are basic hierarchical pictures within the base layer and the blue highlighted boxes are supplementary hierarchical pictures within the enhancement layer. The illustrated enhancement layer corresponds to Demircin's upper layer data and the claimed supplementary video image coding data. When performing decoding with SNR scalability, decoding the "first

enhancement layer requires availability of the base layer,” and decoding “each subsequent enhancement layer requires availability of all previous enhancement layers, as well as the base layer.” Demircin, [0035].

[1.5.2] [a supplementary hierarchical picture] whose coding order and display order are earlier by a factor of a group of pictures including an intra coded picture and a plurality of inter prediction coded pictures than those of a basic hierarchical picture included in the basic video image coding data,

271. The Kimoto-Demircin combination renders element [1.5.2] obvious. First, Kimoto teaches that upper layer data (*supplementary video image coding data*) is not received by the receiving unit until the lower layer data (*basic video image coding data*) has already been received, stored, decoded, and viewed. *E.g.*, Kimoto, [0041]-[0043], [0025]; *supra* [1.3]. The reception of lower layer data sufficient to decode and view a series of pictures in the moving image content at a low quality, prior to the reception of any upper layer data for those pictures, demonstrates that the later received upper layer data includes at least one supplementary hierarchical picture whose coding order and display order is earlier than a basic hierarchical picture in the earlier received and stored lower layer data.

272. Further, in the combination, Kimoto’s upper- and lower-layer data are organized in groups of pictures (GOPs) using a frame structure like that in Demircin’s FIG. 1. *E.g.*, *supra* [1.3], [1.5.1], and *infra* Element [1.5.3]. With a GOP size of just four frames, for example, the received supplementary hierarchical picture in the upper layer would be earlier by a factor of a GOP including an I-frame and a plurality of B-frames

than those of a basic hierarchical picture included in the lower-layer data. Illustration 2 illustrates the point for SNR scalability:

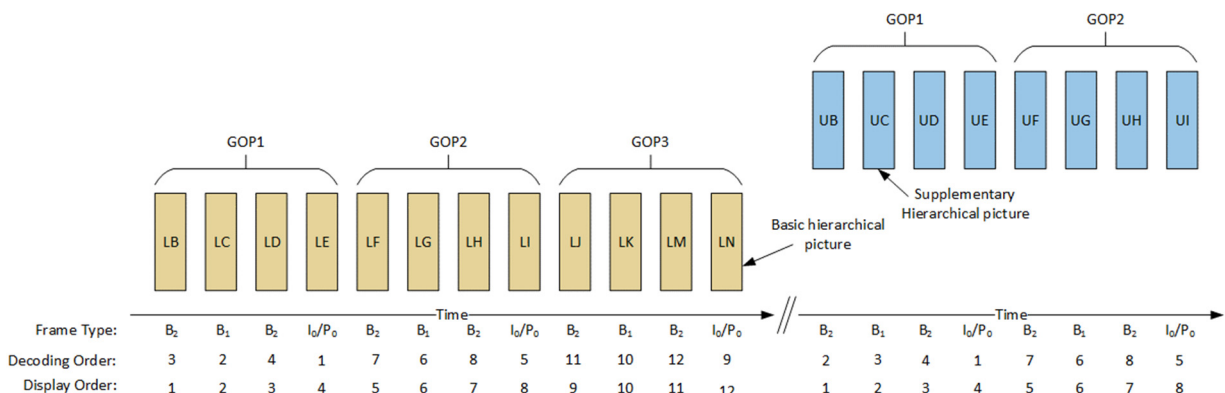


Illustration 2 of Kimoto-Demircin Combination

273. In Illustration 2, lower layer data including basic hierarchical pictures (LB-N) for frames 1-12 are received and stored at an earlier time. *E.g.*, Kimoto, [0041]-[0043]. In the combination, based on Demircin's frame structure, the lower layer data includes GOP1 with intra-coded picture LE and inter-prediction coded pictures LD, LC, and LB; GOP2 with intra-coded picture LI and inter-prediction coded pictures LH, LG, and LF; GOP3 with intra-coded picture LN and inter-prediction coded pictures LM, LK, and LJ.⁷ Later in time (as shown by the broken arrow), based on Kimoto's teachings, the upper layer data for frames corresponding to those in the lower-layer data

⁷ The number of pictures in the lower layer data depends on the size of the video being reproduced, and the number shown in the figure is purely for illustration purposes.

is received (*e.g.*, frames 1-8). *E.g.*, Kimoto, [0067]. In the combination, based on Demircin's frame structure, the upper layer data includes GOP1 with intra-coded picture UE and inter-prediction coded pictures UD, UC, and UB; GOP2 with intra-coded picture UI and inter-prediction coded pictures UH, UG, and UF. Though not shown, the upper layer data could also have included GOP3, depending on the length of the upper layer first data. As taught by Kimoto, however, the upper layer data could be "upper layer first data," which in this example includes GOP1 and GOP2. *See* Kimoto [0067]. Later, subsequent upper layer data is received in Kimoto, which could include GOP3. *See id.*, [0068].

274. As shown by Illustration 2, at least one (but really, multiple) upper layer pictures have a coding order and display order earlier by a factor of a GOP than at least one lower layer picture. For example, UC in GOP1 is a supplemental hierarchical pictures that has a coding order of 3 and a display order of 2 while LN in GOP3 has a coding order of 9 and a display order of 12. GOP1 is separated by GOP3 by GOP2; more particularly, UC in GOP1 is separated from LN in GOP3 by the entirety of GOP2. Thus, UC has a coding order and a display order earlier than LN by a factor of a GOP.

275. In the example, a GOP is equal to one I-frame and three B-frames, for a total of 4 frames. *E.g.*, Demircin, [0037], FIG. 1. An upper layer picture (UC) having a coding order of 3 and a display order of 2 has a coding order and a display order more

than one GOP (and four frames) from a lower layer picture (LN) having a coding order of 9 and a display order of 12.

276. Recall that all of the lower layer data shown in Illustration 2 is received and stored before any of the upper layer data is received. *E.g.*, Kimoto, [0041]-[0043]. Therefore, it would have been obvious that there would be many upper layer pictures having a coding order and display order earlier by a factor of a GOP than a lower layer picture.

277. For each of these reasons, the Kimoto-Demircin combination renders element [1.5.2] obvious.

[1.5.3] a basic hierarchy and a supplementary hierarchy being set in units of the group of pictures;

278. The Kimoto-Demircin combination renders element [1.5.3] obvious. The upper layer data and the lower layer data in the combination are hierarchically coded using an international standard of hierarchical coding, such as H.264/MPEG-4 AVC SVC. *See, e.g.*, Kimoto, [0019]-[0024], [0038]; *see also* Demircin, [0020]-[0021], [0026], [0035]-[0037]; *supra* [1.3]. Demircin's FIG. 1 also discloses hierarchical B-frame coding. Demircin, FIG. 1. As taught in both Kimoto and Demircin, hierarchical coding of the video coding data, including both lower (basic) and upper/enhancement (supplementary) data, can take different forms, including temporal scalability and signal-to-noise ratio (SNR) scalability (also referred to as image quality scalability). *E.g.*, Demircin, [0026]; Kimoto, [0037]. Kimoto teaches that the lower layer data and the

upper layer data are hierarchically coded. *E.g.*, Kimoto, [0019]-[0024]. According to MPEG, a GOP is “the smallest coding unit that can be independently decoded.” SAMSUNG-1015, 54 (emphasis added). In the combination, the lower and upper layer data of Kimoto are likewise set in units of GOPs that can be decoded independently of pictures in other GOPs. SAMSUNG-1003, ¶¶199, 277. According to these teachings, in the combination, the lower layer data and the upper layer data in the combination are set in units of the group of pictures, such as the units of GOP described by Demircin, each unit of which includes an intra-coded picture and a plurality of inter-prediction coded pictures. Demircin, [0036] (“FIG. 1 illustrates an exemplary ... GOP structure and layering structure of a sample SVC bit stream. ... The first frame of the GOP (in encoding/decoding order) is a key picture 2, 4. The key pictures are either intra-coded (I-frame) 2 or inter-coded using a previous (key) picture (P-frame) 4 as a reference for motion-compensated prediction. ... The remaining pictures of a GOP are hierarchically predicted ...”), [0037], FIG. 1.

279. For each of these reasons the Kimoto-Demircin combination renders element [1.5.3] obvious.

[1.6] acquiring basic video image coding data received before supplementary video image coding data that has been received at the moment; and

280. The Kimoto-Demircin combination renders element [1.6] obvious.

281. First, as discussed above (*supra* [1.3] & [1.5.2]), Kimoto teaches that “it is assumed that the low quality content data, or lower layer data, has already been

transferred from a server to the receiving unit 20.” Kimoto, [0025], [0044]. Thus, the lower layer data (*basic video image coding data*) is already stored at the lower data storing unit 202 at the receiving unit before upper layer data (*supplementary video image coding data*) is received at the moment. *Id.* This lower layer data is the basic video image coding data, which includes basic hierarchical picture (or low-quality content data).

282. Then, Kimoto teaches that the “receiving unit 20 receives the upper layer first data 1002 from the transmitting unit 10, and stores the upper layer first data 1002 into the upper layer first data storing unit 200 (S02).” Kimoto, [0067].

283. Next, “[w]hen reproducing the contents, the hierarchical data synthesizing unit 203 synthesizes the upper layer first data 1004 stored in above S03 and the corresponding part 1005 of the lower layer data to generate high quality data 1006 (S04).” Kimoto, [0067]-[0070]. Because the lower layer data is stored in the lower layer data storing unit 202, it would have been obvious that the receiving unit acquires the lower layer data from lower layer data storing unit 202 in order for the hierarchical data synthesizing unit 203 to carry out operations for synthesizing the upper layer data with the lower layer data. Kimoto, [0067]-[0070]. This process of acquiring earlier-received lower layer data (basic video image coding data) from memory is congruent with the understanding of the term “acquiring” from the ’303 Patent specification, which discloses that, “after supplementary hierarchical pictures are received, video image

coding data to be reconstructed from basic hierarchical pictures stored in the memory and the received supplementary hierarchical pictures.” ’303 Patent, 12:6-9, 7:34-43.

284. Kimoto also teaches that reproduction of high-quality images is performed in real time upon receipt of upper layer data after a short buffering interval. Kimoto, [0048], [0053], [0060]-[0064], FIGS. 3-5. “The receiving unit 20 becomes capable of reproducing contents when upper layer first data is stored in the upper layer first data storing unit 200” and “[t]he lower layer data on the reproduction contents is previously stored in the lower layer data storing unit 202 by the point in time when the contents are reproduced.” *Id.*, [0040]-[0041], [0069]-[0071]. Thus, Kimoto discloses and renders obvious acquiring lower layer data from data storing unit 202 for synthesis with upper layer data in the moment the upper layer data is received.

[1.7] reconstructing video image coding data from the basic video image coding data and the supplementary video image coding data.

285. The Kimoto-Demircin combination renders element [1.7] obvious. For example, Kimoto teaches that the receiving unit reconstructs (*e.g.*, synthesizes) the lower layer data (*basic video image coding data*) with the upper layer data (*supplementary video image coding data*) to generate high quality data for reproduction. Kimoto, [0019], [0025], [0042], [0069]-[0071], [0092]-[0097], FIG. 2 (S05-S06). Noteworthy is that the ’303 Patent also “*synthesizes* basic video image coding data and supplementary image coding data” using a “coding data reconstruction unit.” ’303 Patent, 12:53-55 (emphasis added).

Kimoto, therefore, clearly teaches reconstructing video image coding data as recited in claim 1.

X. Additional Remarks

286. I currently hold the opinions set expressed in this declaration. But my analysis may continue, and I may acquire additional information and/or attain supplemental insights that may result in added observations.

EXHIBIT A

Cliff Reader, Ph.D.

Expertise

- Real-time Processing and Display
 - Video coding (MPEG-HEVC)
 - Audio coding (MPEG-AAC)
 - Imaging/Video Systems Architecture
 - Imaging/Video Chip Architecture
 - Videoconferencing
 - Speech Processing
 - Consumer audio/video
 - Reconnaissance Imaging
 - Medical Imaging
-

Professional Summary

Dr. Reader has over forty five years of work experience in digital imaging, digital video & audio and digital speech processing. In the 1970's and early 1980's, he was one of the leaders in the digital imaging field, performing extensive work in areas of compression, real-time processing, real-time display, image enhancement, and others. Applications included reconnaissance imaging, medical imaging, earth resources management and videoconferencing. This work anticipated much of the technology being used in contemporary consumer digital video systems. He has in-depth experience of the commercial history and when key parts of technology were first developed and published.

Dr. Reader's career includes technical work in areas of algorithm design, system design, and semiconductor chip design; and business development, marketing and sales work in areas of market analysis, product roadmap development and direct sales.

Dr. Reader has held many leadership roles in standards development: US Head of Delegation to MPEG (1991-1993); chief editor of the MPEG1 standard (1991-1993); instigator of the successful effort to make MPEG2 the market standard for digital TV over proprietary competitors in the US and national competitors in Europe and Japan (1992); MPEG4 Subcommittee Chairman (1993-1996); AVS (China) Subgroup Chairman (2003-Present). He is an expert in all major audio-video standards. MPEG audio (Layer I, II, mp3, AAC), Dolby AC3, MPEG video (1, 2, 4), H.264, Windows Media 9 (VC-1), H.265 (HEVC), AVS and AV1. He is also an expert in DVD standards, including formatting, navigation and authoring.

Dr. Reader has extensive experience in patent licensing and patent pools. He was the technical expert for developing the initial list of potentially essential patents for the first MPEGLA patent pool (MPEG2 video and systems). He is the Co-Director leading the pool formation and license negotiations for the AVS patent pools. Dr. Reader has experience in the valuation of patents and the acquisition and sale of patent portfolios.

Dr. Reader has authored over twenty technical papers, two book chapters and two patents in digital video processing, display, and compression. He taught digital signal processing and image processing at Santa Clara University in the early 1980s and is an Adjunct Professor at Peking University, Beijing, China.

Employment History

From: 2001 **Independent Consultant**
To: Present Escondido, CA
Provide technical and marketing consulting services in the areas of digital imaging, digital video, digital audio and speech including:

- consumer video
- real-time processing and display
- image and video & audio compression (specializing in standards)
- imaging/video systems architecture
- imaging/video chip architecture
- audio/video transmission, including optical & wireless networks

Over 80 clients including Apple, AT&T, Adobe, Broadcom, China Central TV, Chinese Academy of Sciences, Cisco, Comcast, DirecTV, Dish Network, Disney, EchoStar, ETRI, Facebook, Fujifilm, Gateway, GE, Google, Haier, Hisense, HP, Hulu, IBM, Intel, Johnson Controls, KAIST, Lenovo, LG, Matsushita, MediaTek, Microsoft, Motorola, Netflix, Nikon, Nokia, ON2, OPPO, Peloton, Philips, Quanta, Riverbed, Samsung, Sharp, Siemens, Sony, Starz, Sun Microsystems, Toshiba, TPV, Tyco, US Department of Justice, Verizon, Vimeo, Vizio, Wistron, Xiaomi, Yahoo, Zoran and ZTE.

Personal database of patent analysis:

- Evaluation of all the US patents in the MPEG2, H.264 and H.265 patent pools.
- Evaluation of the US patents and published applications underlying the H.264 and H.265 standards.
- Personal archive of over 2000 audiovisual patents.

Personal database of audio and video prior art.

- Virtually complete collection of all contributions to the H.26x and MPEGx families of standards from 1985 to today.
- Many hundreds of technical journal and conference articles documenting the first invention of key technologies.
- Work-in-progress book on the history of video compression.
- Historical collection of technical data from the earliest digital video, image processing and real-time image display systems in the 1970s and 1980s.

From: 1999 **nDSP Corporation**
To: 2001 Campbell, CA
Title: *Vice President of Marketing and Acting Vice President of Sales*

Responsible for marketing and sales in a startup company providing high-quality video signal processors to TV manufacturers. Developed product specifications and roadmap. Built and managed sales and marketing team of 6 direct and contractor staff in the US and Japan to provide technical sales, sales support, application engineering, Marcomm and PR. Established distribution in China, Japan and throughout the Far East. Secured first sales as company moved into production phase with design wins in China, Japan and Europe. Initiated and secured major design wins in emerging Progressive DVD business, with production volumes beginning in 4Q00 at an annualized rate of \$3.5M. Established strategic partnership with Pixelworks in LCD chip business. Pixelworks acquired nDSP in 2001.

From: 1997 **CagEnt Technologies, Inc. (wholly owned by Samsung)**
 To: 1998 Santa Clara, CA
 Position: *Vice President of Sales, Digital Video Products*
 Responsible for a line of PC-based real-time MPEG2 and Dolby Digital encoders. Established distribution agreements for third-party DVD authoring software. Secured OEM contracts for the products. Built the sales channel worldwide. Built sales to break-even point in 9 months leading to sale of the product line.

From: 1993 **Samsung Semiconductor, Inc.**
 To: 1997 San Jose, CA
 Position: *Director of Marketing, Digital Media Products;*
Acting Director of Software Engineering
 Developed business plans for digital audiovisual semiconductor chips. Managed team of over 50 software engineers in US and Korea for development of a media processor. Established strategic partnerships with Microsoft and IBM for media processor

From: 1990 **Cypress Semiconductor, Inc.**
 To: 1993 San Jose, CA
 Position: *Business Development Manager*
 Founding member of new business unit chartered to develop a “video DSP”. Jointly developed a complete business plan, and personally initiated and developed relationships with Philips, Sun Microsystems, Apple Computers and IBM. Chief architect for this early media DSP design, capable of complete MPEG1 decode in real time. (Fully validated by cycle-accurate simulation). Managed R&D team of 10 hardware and software engineers.

Head of US Delegation to MPEG, 1991 to 1992. Chief Editor of the MPEG1 standard. Technical Expert for establishing the

MPEGLA Patent Licensing Pool.

From: 1988 **Sun Microsystems, Inc.**
To: 1990 Mountain View, CA
Position: *Program Manager*
Developed requirements and specifications for real-time imaging system. Managed a team of 20 engineers to develop the architecture and high-level design. Interfaced with two key customers – Philips Medical Systems and Lockheed. Developed design for single-board real-time medical and reconnaissance image display.

From: 1981 **International Imaging Systems, Zoran, Virtual Imaging**
To: 1988 Design and development of real-time image processing and display systems. Details available on request.

From: 1975 **Ford Aerospace & ESL (A Division of TRW)**
To: 1981 Details available on request.

Consulting History

Date 2023 **Intel**
To: Present
Duties: Consulting Expert

Date 2023 **Disney**
To: Present
Duties: Consulting Expert

Date 2023 **Samsung**
To: Present
Duties: Consultant

Date 2023 **Peloton**
To: Present
Duties: Consulting Expert in NEC matter

Date 2023 **Alps Alpine**
To: 2023
Duties: Consulting Expert in DivX matter

Date 2023 **Fish & Richardson**
To: Present
Duties: Consultant

Date	2023	Oppo
To:	Present	
Duties:	Consulting Expert on speech processing in Nokia matter	
Date	2022	Samsung
To:	2023	
Duties:	Testifying Expert on speech processing in Jawbone matter	
Date	2022	Asus, Lenovo, MSI
To:	2023	
Duties:	Consulting Expert in VideoLabs matter	
Date	2022	Google
To:	2023	
Duties:	Consulting Expert on speech processing in Jawbone matter	
Date	2022	Starz
To:	Present	
Duties:	Consulting Expert in VideoLabs matter	
Date	2022	Meta Platforms (Facebook)
To:	Present	
Duties:	Consulting Expert in VideShare matter	
Date	2022	Oppo
To:	Present	
Duties:	Consulting Expert in InterDigital matter	
Date	2022	Apple
To:	2023	
Duties:	Consulting Expert on speech processing in Jawbone matter	
Date	2021	Verkada
To:	2022	
Duties:	Testifying Expert in Motorola Solutions matter	
Date	2021	Facebook
To:	2022	
Duties:	Consulting Expert in EyesMatch matter	
Date	2021	Videoconferencing company
To:	2023	
Duties:	Testifying Expert	
Date	2021	Canon

To:	2022	
Duties:		Consulting Expert in WSOU Investments/Brazos Licensing matter.
Date	2021	Axis
Duties:		Consulting Expert
Date	2021	nVIDIA
To:	2022	
Duties:		Consultant
Date	2020	Xiaomi
To:	2021	
Duties:		Consulting Expert in InterDigital matter
Date	2020	LG
To:	2021	
Duties:		Consulting Expert DivX matter
Date	2020	Vizio
To:	2023	
Duties:		Consulting Expert VDPP matter
Date	2020	Netflix
To:	2022	
Duties:		Consultant
Date	2020	Oppo
To:	Present	
Duties:		Consultant
Date	2020	IdeaHub
To:	Present	
Duties:		Preparation of declaration for Inter Partes Review proceedings. (Unified Patents matter)
Date	2019	Google
To:	2021	
Duties:		Preparation of declarations for Inter Partes Review proceedings (PMC matter)
Date	2019	Google
To:	2021	
Duties:		Consulting Expert (Multiple projects)
Date	2019	LG
To:	2020	

	Duties:	Preparation of declarations for Inter Partes Review proceedings. (Uniloc matter)
Date	2019	Klarquist, Sparkman
To:	2021	
	Duties:	Consulting Expert (Multiple projects)
Date	2019	Netflix, Hulu
To:	2021	
	Duties:	Preparation of declarations for Inter Partes Review proceedings. (DivX matter)
Date	2019	MediaTek
	Duties:	Consulting Expert in American Patents matter
Date	2019	AT&T
	Duties:	Consulting Expert
Date	2019	KAIST/KBS
To:	2021	
	Duties:	Consulting Expert for Inter Partes Review proceedings. (Unified Patents matter)
Date	2019	Vimeo
To:	2020	
	Duties:	Preparation of declarations for Inter Partes Review proceedings. (BT matter)
Date	2019	Apple
	Duties:	Preparation of declarations for Inter Partes Review proceedings. (Red.com matter)
Date	2019	Adobe
To:	2021	
	Duties:	Preparation of declarations for Inter Partes Review proceedings and consulting expert. (Realtime matter)
Date	2018	Toshiba
To:	2019	
	Duties:	Testifying expert concerning valuation of essential patents for ATSC standards
Date	2018	Hisense
	Duties:	Analysis of essential patents for AAC standards
Date	2018	HTC

To:	2020	
Duties:		Analysis of essential patents for the H.265 standard
Date	2018	Cisco
To:	2020	
Duties:		Consulting expert in Realtime matter
Date	2017	Dish Network
To:	2023	
Duties:		Consulting expert in ClearPlay matter
Date	2017	Apple
To:	2019	
Duties:		Preparation of declarations for Inter Partes Review proceedings. (Uniloc matter)
Date	2017	Dish Network
To:	2020	
Duties:		Testifying expert in Realtime matter
Date	2017	CustomPlay
To:	2020	
Duties:		Preparation of declaration for summary judgment in Amazon case.
Date	2017	Vizio, Funai, LG, MediaTek, MStar
Duties:		Testifying expert in Broadcom case.
Date	2017	Apple
Duties:		Preparation of declarations for Inter Partes Review proceedings. (Nokia matter)
Date	2017	Vizio
Duties:		Preparation of declarations for Inter Partes Review proceeding. (Lightside Technologies matter)
Date	2016	Dell, Riverbed, EchoStar, Dish Network
To:	2019	
Duties:		Testifying expert in Realtime matter
Date	2016	Sharp, Vizio, ON
To:	2018	
Duties:		Testifying expert in WiLAN matter
Date	2016	Avid
To:	2017	
Duties:		Testifying expert in Harmonic matter.

Date	2016	Apple
Duties:		Testifying expert in PUMA matter.
Date	2016	Vizio
Duties:		Testifying expert in PMC matter.
Date	2016	US Dept. of Justice, Dept. of Homeland Security, DoD, FBI, NSA, & Others
To:	2021	
Duties:		Consulting expert in Discovery Patents (3 rd Eye Surveillance) matter. Declarations in support of six Inter Partes Reviews.
Date	2015	ST Micro
To:	2016	
Duties:		Consulting expert in Avago matter.
Date	2015	AT&T
Duties:		Consulting expert in Raniere matter.
Date	2015	Motorola
To:	2016	
Duties:		Consulting expert in PUMA matter.
Date	2015	TPV
To:	2016	
Duties:		Consulting expert in Thomson licensing matter.
Date	2015	Tyco (Exacq, Johnson Controls)
To:	2019	
Duties:		Consulting expert in JDS Technologies matter.
Date	2015	Tyco
Duties:		Declaration in support of Inter Partes Review petition in Hawk Technology matter.
Date	2015	Sceptre
Duties:		Expert reports in licensing matter concerning MPEGLA MPEG2 pool patents.
Date	2015	Samsung
Duties:		Consultant and fact witness in case asserting Samsung patents of which I am a named inventor, against NVidia.
Date	2015	Nissim
To:	2016	

	Duties:	Consultant expert for patent holder in 20 th Century Fox, MGM, and Paramount cases concerning patents related to DVD navigation. Testified at Markman hearings.
Date	2015	Nikon
To:	Present	
	Duties:	Consultant.
Date	2015	Humax
	Duties:	Search and analysis of prior art.
Date	2015	Apple
	Duties:	Analysis of patent portfolio potentially essential for the H.264 standard.
Date	2014	Haier
To:	2018	
	Duties:	Analysis of patent portfolios for smart consumer devices.
Date	2014	Cisco
To:	2021	
	Duties:	Consultant.
Date	2014	ZTE, Sony, LG, Samsung, Nokia, GoPro
To:	2015	
	Duties:	Consulting expert.
Date	2014	Tyco
To:	2015	
	Duties:	Expert reports; declaration in support of Inter Partes Review in Trover matter.
Date	2013	Humax
	Duties:	Consultant.
Date	2014	GE Licensing
To:	2015	
	Duties:	Consultant.
Date	2014	HikVision & Axis
To:	2015	
	Duties:	Declaration in support of Inter Partes Review. Expert reports in Trover matter.
Date	2014	Sharp
To:	2016	
	Duties:	Declaration in support of Inter Partes Review in Wi-Lan matter.

Date	2013	InterDigital
To:	2015	
Duties:	Consulting expert in support of licensing a patent portfolio.	
Date	2013	Apple
Duties:	Consulting expert in National Cheng Kung University matter.	
Date	2013	Acer, ASUS, TPV
To:	2014	
Duties:	Testifying expert in Florida Atlantic University matter.	
Date	2013	Toshiba
To:	2014	
Duties:	Testifying expert in WiLAN matter.	
Date	2013	Grass Valley Group
Duties:	Consulting expert for defense in FastVDO matter.	
Date	2013	TI
To:	Present	
Duties:	Consulting expert.	
Date	2012	Vizio
Duties:	Consulting expert in Innovus Prime matter.	
Date	2012	Haier & Shandong Provincial Government
To:	2014	
Duties:	Foreign expert advisor for developing and acquiring patent portfolios for innovative consumer products. Development of a patent pool for patents essential to standards for wireless power.	
Date	2012	Vizio
To:	2014	
Duties:	Consulting expert in MIT matter.	
Date	2012	Nokia
To:	2013	
Duties:	Consulting expert.	
Date	2012	TPV & Vizio
To:	2013	
Duties:	Testifying expert for defense in Hitachi matter.	
Date	2012	Samsung
Duties:	Consulting expert in Myport matter.	

Date	2012	Verizon Wireless (Cellco) & T-Mobile
To:	2013	
Duties:		Testifying expert for defense in Realtime Data matter.
Date	2011	Apple
Duties:		Consulting expert in Technicolor matter.
Date	2012	Microsoft
To:	2014	
Duties:		Consulting expert for plaintiff in Motorola matter.
Date	2012	Apple & LG
To:	2014	
Duties:		Testifying expert for defense in MPT matter.
Date	2011	Sanyo & Olympus
Duties:		Consulting expert for defense in FastVDO matter.
Date	2011	Apple
Duties:		Consulting expert for defense in FastVDO matter.
Date	2011	Sony
Duties:		Consulting expert for defense in FastVDO matter.
Date	2011	Philips
To:	2014	
Duties:		Consulting expert for licensing essential patents.
Date	2011	Apple
Duties:		Consulting expert for defense in MobileMediaIdeas matter.
Date	2011	Video264Innovations
Duties:		Consulting expert.
Date	2011	Nikon, Fujifilm & Panasonic
Duties:		Consulting expert for defense in FastVDO matter.
Date	2011	Haier
To:	2014	
Duties:		Consultant for standards, patent pools and patent licensing.
Date	2011	Susman Godfrey
Duties:		Consultant for patent analysis.
Date	2011	Vizio

	Duties:	Testifying expert for defense in MPT matter.
Date	2011	Tivo
	Duties:	Consulting expert.
Date	2011	Nikon
To:	2013	
	Duties:	Consulting expert for defense in MPT matter.
Date	2011	Broadcom
	Duties:	Consulting expert.
Date	2010	Lionsgate Entertainment
	Duties:	Consulting expert in MPEG licensing matter.
Date	2010	Verizon
To:	2012	
	Duties:	Testifying expert for defense in ActiveVideo Networks matter.
Date	2010	Lenovo
To:	2011	
	Duties:	Consulting expert for defense in licensor matter.
Date	2010	Sonic, Cyberlink & Joint Defense Team
	Duties:	Consulting expert for defense in Mediostream licensor matter.
Date	2010	Wistron
To:	2011	
	Duties:	Consulting expert for defense in Toshiba licensor matter.
Date	2010	Panasonic, Pioneer
To:	2011	
	Duties:	Consulting expert for defense in VES licensor matter.
Date	2010	Zoran
To:	2011	
	Duties:	Technical analysis of patents potentially related to video coding standards.
Date	2010	Intravidual
To:	2015	
	Duties:	Consulting expert for patent licensing.
Date	2009	Apple
	Duties:	Technical analysis of patents potentially related to video coding standards.

Date	2009	SanDisk
Duties:		Consulting expert re Audio and Video coding patents
Date	2009	Lenovo
To:	2011	
Duties:		Consulting expert for defense in MPEGLA licensor matter.
Date	2009	Polycom
To:	2010	
Duties:		Consulting expert for defense in MPT matter.
Date	2009	DirecTV
To:	2011	
Duties:		Consulting expert for defense in MPT matter.
Date	2009	EchoStar
To:	2010	
Duties:		Consulting expert for defense in MPT matter.
Date	2009	Vizio
Duties:		Consulting expert for defense in Sony matter.
Date	2009	Vizio
To:	2010	
Duties:		Consulting expert for defense in LG matter.
Date	2009	Vizio, TPV, Amtran
To:	2010	
Duties:		Technical report & video presentation to US Customs on design issues related to Funai matter. Testified in ITC hearing, March 2010.
Date	2008	Humax (Korea)
To:	2009	
Duties:		Audio patent and prior art review.
Date	2008	IC Media (Canada)
To:	2009	
Duties:		Design of digital signage.
Date	2008	AVS Patent Pool Administration (China)
To:	Present	
Duties:		Co-Director, leading license negotiations among domestic and multinational licensors.

Date	2008	DirecTV
Duties:		Prior art search for DirecTV's appeal against Finisar.
Date	2008	Pure Digital
Duties:		Prior art research and analysis.
Date	2008	Wragge (UK)
Duties:		Prior art review for two video coding patents.
Date	2008	Microsoft
Duties:		Extensive non-infringement analysis for patents asserted by Multimedia Patent Trust (MPT).
Date	2008	iPotential
To:	2010	
Duties:		Technical analysis of patents potentially related to video coding standards.
Date	2008	Alba (UK) aka Harvard Research
To:	2009	
Duties:		General education on essential patents, licensing, prior art related to the MPEG2 video standard.
Date	2008	Philips (Holland)
Duties:		Review of patents for essentiality.
Date	2008	Proview
Duties:		Prior art search for TV display features.
Date	2008	Nissim
To:	2014	
Duties:		Consultant expert for plaintiff in Time-Warner case concerning patents related to DVD navigation.
Date	2008	Comcast
Duties:		Technical analysis of MPEG2 systems, DSM-CC, and carousel operation in the cable TV network environment. Contribution to expert's report in Comcast's successful litigation against Finisar.
Date	2007	Sun Microsystems
To:	2009	
Duties:		Chief architect of royalty-free video codec: OMS Video. Extensive supporting patent analysis and technical publication analysis for prior art review.
Date	2008	SanDisk

	Duties:	Technical analysis of patents and prior art searching for audio coding.
Date	2008	Stamberger (Munich)
	Duties:	Technical analysis of patents potentially related to video coding standards.
Date	2008	Trigem America Liquidating Trust
	Duties:	Preparation of expert reports concerning litigation between Toshiba and Trigem.
Date	2007	Susman-Godfrey
	Duties:	Technical analysis of patents potentially related to video coding standards.
Date:	2007	Wong-Cabello
	Duties:	Technical analysis of patents potentially related to video coding standards.
Date:	2007	Sisvel
	Duties:	Technical assistance in licensing audio patents to a third party.
Date:	2007	Quanta
	Duties:	Technical assistance in litigation related to DVD.
Date:	2006	CCTV (China Central TV)
	Duties:	Technical analysis of patents potentially related to MPEG2, H.264, AVS and VC-1 video coding standards and MPEG-AAC and Dolby audio coding standards.
Date:	2006	MediaTek
	Duties:	Technical analysis of MediaTek patents potentially related to video coding standards.
Date:	2006	Alcatel
	Duties:	Technical analysis of Alcatel patents potentially related to video coding standards.
Date:	2006	Motorola
	Duties:	Analysis of technologies contributed to the H.264 standard.
Date:	2005	Nokia
	Duties:	Analysis of technologies contributed to the H.264 standard.
Date:	2005	Avid
	Duties:	Technical analysis of patents potentially related to coding standards.

Date:	2005	IBM
Duties:		Analysis of technologies contributed to the H.264 standard.
Date:	2005	HP
Duties:		Analysis of selected audio coding patents.
Date:	2005	ComVentures
Duties:		Due diligence review of a start-up company in the video semiconductor business.
Date:	2005	nVidia
Duties:		Prior art search for a set of third-party patents.
Date:	2004	Microsoft
Duties:		- Working with a major financial services company, developed a procurement model for managing licensing costs. - Assisting development of non-infringement positions for a set of patents. (Not in suit).
From:	2004	SAGEM (France)
To:	2005	
Duties:		Technical analysis of potential infringement by SAGEM of a set of third-party patents.
From:	2004	Polycom Inc.
To:	2005	
Duties:		Technical analysis of Polycom patents potentially related to video coding standards.
From:	2004	Sony Corp.
To:	2006	
Duties:		- Technical analysis of potential infringement by Sony of a set of third-party patents. The parties settled without litigation. - Analysis of a third-party patent per Sony's direction.
Date:	2004	On2 Inc
Duties:		Delivery of the comprehensive set of reports on IPR potentially related to the emerging H.264 standard.
Date:	2004	Matsushita Avionics
Duties:		Technical analysis of potential infringement of Matsushita patents by a third party.
Date:	2004	Connex Inc.
Duties:		Teaching the details of key complex algorithms in the H.264 standard.

Date:	2004	DoCoMo (Japan)
Duties:		Delivery of the comprehensive set of reports on IPR potentially related to the emerging H.264 standard.
From:	2004	DivX Inc.
To:	2006	
Duties:		Analysis of technologies contributed to the H.264 standard
Date:	2004	Cues, Inc
Duties:		Assisted evaluation of infringement by Cues' products of a patent asserted by a third party. The parties settled without litigation.
From:	2003	GE Licensing
To:	2004	
Duties:		Working with a major financial services company, developed a business, market and financial model for valuation of a portfolio of patents through their expiration.
From:	2003	Institute of Computing Technology, Chinese Academy of Sciences
To:	Present	
Duties:		• Preparation of IPR Report for IPR potentially related to the AVS Video standard. This effort involved review of over 2500 abstracts and over 450 patents. • IPR Subgroup Chairman of AVS. Formed an IPR Experts group to establish a legal foundation for AVS IPR. This group comprised approximately 20 attorneys and business development managers from eight multinational companies representing the PC, consumer, semiconductor and communication industries. (Broadcom, IBM, Intel, Matsushita, Microsoft, Nokia, Sony, Sun Microsystems). The group drafted a member agreement, bylaws and a formal IPR policy governing licensing commitments, disclosure obligations, and protection of license holder rights. AVS Members approved these documents in September 2004. • Co-Director of AVS Patent Pool. Leading the drafting and negotiation of patent pool management and licensing agreements. • Adjunct Professor, supervising a graduate student project on deinterlacing algorithm development.
Date:	2003	Yahoo
Duties:		Prior art search to provide background information per Yahoo's request.
Date:	2002	Gateway, Inc.
To:	2007	
Duties:		Extensive prior art search and assistance with claim construction in a litigation matter with Lucent Technologies.

From:	2002	Apple Computer
To:	2007	
Duties:		• Technical analysis of Apple patents potentially related to H.264. • Prior art search for various patents per Apple's request.
Date:	2002	Siemens (Germany)
Duties:		• Technical analysis of Siemens patents potentially related to H.264. • Delivery of the comprehensive set of reports on IPR potentially related to the emerging H.264 standard.
Date:	2002	JVT Committee (Apple, Cisco, Nokia, Polycom, Sun Microsystems, Tandberg, Teles, Texas Instruments, UBVideo, VideoLocus)
Duties:		Developed a comprehensive set of reports on IPR potentially related to the emerging H.264 standard. The reports included a technical review of legacy and new patents for video coding, and analysis of the patent statements provided by all companies whose contributions were adopted into H.264.
Date	2002	Intervideo
Duties:		Advised on market opportunities in the emerging MPEG4 marketplace.
Date:	2001	Proskauer Rose LLP
Duties:		Provided background support in MPEGLA's (successful) effort to license MPEG2 to the major PC manufacturers.
Date	2001	Pixonics
Duties:		Developed marketing plan and marketing support materials for this incubator-stage company.
Date	2001	Eltron, (Taiwan)
Duties:		Provided overview of the digital video semiconductor market.
Date	1999	nDSP Corporation
Duties:		Drafted the sales and marketing plan for this early-stage company. Joined nDSP full-time.
Date:	1998	IBM Corp.
Duties:		Assisted with participation in the ATSC DASE (datacasting) standard.
Date	1998	Spruce Technologies
Duties:		Assisted with transition of Cagent MPEG encoder product after it was sold to Spruce.
From:	1997	Samsung Advanced Institute of Technology
To:	1999	

Duties: Supported Samsung researchers in MPEG4 standard development. Led effort to provide error-resilient 3D-mesh coding.

Date 1997 **MPEGLA LLC**
Duties: Review of opportunity for an MPEG4 patent pool

From: 1993 **CableLabs**
To: 1994
Duties: Teamed with a patent attorney to develop the initial essential patent list for MPEG2. Reviewed approximately 10,000 abstracts and 1,000 patents for the MPEG2 Video and MPEG2 Systems standards. This effort led to the MPEGLA patent pool.

Litigation Support Experience

Quinn Emanuel Urquhart & Sullivan, LLP, San Francisco, CA
Case Motorola Solutions vs. Verkada, Investigation No. 337-TA-1281
Project: Expert report, deposition and trial testimony on invalidity of Motorola patents.
Status: Final determination found Verkada not infringed
Filed: August, 2021
Location: United States International Trade Commission, Washington DC

Orrick, Herrington & Sutcliffe, LLP, New York, NY
Case ClearPlay vs. Dish Network & EchoStar Technologies, Case No. : 2:14-cv-00191
Project: Expert report, deposition and trial testimony on invalidity of ClearPlay patents.
Status: ClearPlay patents not infringed
Filed: March 13, 2014
Location: United States District Court for the District of Utah

Akin Gump, Fish & Richardson, Finnegan, Baker Hostetler, Washington, DC
Case Broadcom vs. Vizio, LG, Funai, MediaTek, MStar, Investigation No. 337-TA-1047
Project: Expert report and deposition on non-infringement of Broadcom's patent.
Status: Final Determination in favor of Respondants.

Jones Day, Cleveland, OH
Case Avid Technology vs. Harmonic, Case No. C.A. No. 12-627-GMS
Project: Expert reports and deposition on infringement of Avid's patent.

Status: Parties settled before trial.

Haynes and Boone, Dallas, TX
Case Realtime vs. Riverbed, Case No. 6:1S-CV-463-RWS-JDL
Project: Expert report, deposition and trial testimony (E. District Texas) on
invalidity of Realtime patent.
Status: Jury found Realtime patent invalid at trial.

Akin Gump, LLP, Irvine, CA
Case Personalized Media Communications “PMC” vs. TPV et al, Case No.
2:15-cv-01366-JRG-RSP
Project: Declaration in support of claim construction brief.
Status: Parties settled.

Feinberg, Day, LLP, Menlo Park, CA
Case Parthenon Unified Memory Architecture “PUMA” vs. Apple, Case No.
9:12-cv-80701
Project: Expert report and deposition concerning unified memory architecture.
Status: Trial stayed at request of PUMA. Parties settled.

Carey, Rodriguez, LLP, Miami
Case *Twentieth Century Fox Home Entertainment, LLC, Paramount Pictures
Corp., vs. Nissim Corp., Case No. 14-81349/813450-cv-MARRA
Project: Expert witness for patent owner in counterclaims for infringement.
Declaration and expert testimony at Markman Hearings concerning DVD
standards essential patents.
Status: Parties settled before trial.

*Nissim asserted its patents against Twentieth Century Fox and
Paramount. The latter filed a complaint seeking a declaratory judgment
of non-infringement and invalidity.

**TechKnowledge Law Group, Turner Boyd LLP & O’Melveny &
Myers LLP, Los Angeles,**
Case Florida Atlantic University Research Corporation vs. Acer, ASUS &
TPV, Case No. 9:12-cv-80701
Project: Expert testimony at Markman Hearings concerning scan converter
design. Expert reports and deposition testimony.
Status: Parties settled before trial.

Finnegan, Henderson, Farabow, Garrett & Dunner, LLP
Case WiLAN vs. Toshiba, Case No. 1:12-Civ-23744
Project: Expert reports & deposition testimony concerning digital television
design.
Status: Parties settled before trial.

O’Melveny & Myers LLP, Los Angeles & Akin Gump LLP, Dallas

Case Hitachi vs. TPV, Case No. 2:10-CV-260-JRG
Project: Expert reports, depositions & trial testimony (E. District Texas) concerning digital television design.
Status: Jury found no infringement of Hitachi's patents by TPV and found Hitachi's patents invalid. (Partially reversed by JMOL)

Fish & Richardson P.C. San Diego
Case Multimedia Patent Trust vs. Apple & LG, Case No. 10-cv-2618 H
Project: Expert report, & deposition testimony concerning development of technology, standards and patents for video coding.
Status: Jury found no infringement of MPT's patents and found MPT's patents invalid.

Gibson Dunn LLP, Palo Alto
Case Realtime Data vs. Verizon Wireless & T-Mobile, Case No. 6:10-cv-00493-LED
Project: Expert reports, depositions and testifying expert in jury trial (E. District Texas) concerning infringement and validity of patents related to data transmission.
Status: Jury found no infringement of Realtime Data's patents by T-Mobile, and found Realtime Data's patents invalid.

O'Melveny & Myers LLP, San Francisco
Case MobileMediaIdeas vs. Apple, Case No. 10-258-SLR
Project: Expert reports and deposition, concerning infringement and validity of patents related to MPEG Audio Layer III ("mp3") coding.
Status: Trial bifurcated.

Jones Day, Cleveland
Case Multimedia Patent Trust Vs. Vizio, Case No. 09-CV-0278-H (CAB)
Project: Expert report, concerning infringement of patents related to H.264 video coding. (Slated for deposition and jury trial)
Status: Parties settled on the day before my deposition.

Kellog, Huber, et al, Washington, DC
Case Active Video Vs. Verizon, No. 2:10-cv-00248-RAJ-FBS; 2011
Project: Expert report, deposition and testifying expert in jury trial concerning infringement of patents related to video-on-demand
Status: Jury finding of infringement with significantly reduced damages and no finding of willful infringement.

Jones Day, Washington, DC
Case Funai Vs. Vizio, et al, ITC Investigation No. 337-TA-617; 2010
Project: Expert report, deposition and testifying expert in ITC hearing concerning infringement of a patent related to ATSC broadcast standards
Status: Parties settled after the hearing following reports adverse to complainant.

Wilson Sonsini, Goodrich and Rosati, Palo Alto, CA.

Case Terayon vs. RGB Networks; 2007
Project: Expert report for defendant in a contract dispute alleging violation of a non-compete agreement.
Status: Mediation hearing resolved the dispute to the satisfaction of both parties.

Wilmer Cutler Pickering Hale and Dorr, Boston, MA

Case Qualcomm Inc. vs. Broadcom Corp. No. 05 CV 1958 B (BLM); 2007
Project: Expert report and deposition on behalf of Broadcom against assertions of infringement of patents concerning video encoding and decoding, and on Qualcomm's compliance with the policies and procedures for IPR policy in the H.264 standards committee (JVT of ITU-ISO/IEC).
Status: Broadcom found not guilty of infringing Qualcomm's patents. The patents were determined to be not essential for implementing the H.264 standard. Qualcomm found guilty of inequitable behavior in the manner of their participation in the H.264 standards committee. Qualcomm's patents held unenforceable against H.264.

Greenberg Glusker Fields Claman Machtinger & Kinsella.

Case DTS vs. MediaTek; 2005
Project: Expert report for defendant in a contract dispute related to hacking into semiconductor chips to enable functionality for which a license fee was payable.
Status: Mediation hearing resolved the dispute to the satisfaction of both parties. The Mediator personally thanked me for the clarity of my report, which covered technical, business and licensing issues.

Arvay Finley Barristers, Victoria, BC, Canada.

Case AVT Audio Visual Telecommunications vs. University of British Columbia; 2005
Project: Expert report and testimony on behalf of AVT in a contract dispute related to mobile video communication including failure to disclose intellectual property.
Status: Arbitration hearing resulted in adverse ruling to AVT.

Thelen Reid & Priest, San Jose, CA.

Case Toshiba Corp. vs. Trigem Computer Inc. & Trigem America Corp. Case No. CV-03-4558-GHK(AJWx); 2005
Project: Expert report on behalf of Trigem regarding invalidity of patents relating to video and sub-picture data in multimedia computer systems.
Status: Suspended due to insolvency of Trigem.

Morrison & Foerster LLP, San Francisco, CA.

Case	<u>Tivo Inc. vs. Echostar Communications Corp., Echostar DBS Corp., Echostar Technologies Corp., & EchoStar LLC, Case No. 2-04CV-01; 2005</u>
Project:	Expert report and deposition on behalf of Echostar against assertions of infringement of a patent concerning personal video recorders (PVR).
Status:	Adverse ruling to Echostar.
	Pinsent Curtis Biddle, Leeds, UK
Case	<u>Thomson Licensing S.A. and Deutsche Thomson Brandt GmbH vs. Pace Micro Technology PLC. Case No. HC 2000 No. 1584; 2001</u>
Project:	Expert report and slated to testify on behalf of Pace arguing invalidity of patent and defending against assertions infringement of a patent concerning interlaced video coding.
Status:	Parties agreed to settlement on the second day of trial. Settlement terms were very favorable to defendant, and shortly after the trial, Thomson joined the MPEGLA patent pool.

Education

1974	University of Sussex, England	Ph.D., Applied Sciences (EE/CS) Thesis on “Orthogonal Transform Coding of Still and Moving Pictures”. All research performed in residence at the Image Processing Institute, USC Los Angeles, CA.
1970	University of Liverpool, England	Bachelor of Engineering with Honours, Physical Electronics

Publications (Partial List)

C Reader, “Orthogonal Transform Coding of Still and Moving Pictures”, Ph.D. thesis, U. Sussex, England, 1973.

C. Reader, “Intraframe and Interframe Adaptive Transform Coding.” SPIE Vol. 66, pp. 108-117, *Efficient Transmission of Pictorial Information* (1975).

C. Reader, W.H. Chen, “Local Retransmission of Compressed Imagery Data”, SPIE Vol. 149 pp. 196-204, *Applications of Digital Image Processing* (1978).

L. Hubble, C. Reader, “State of the Art in Image Display Systems”, SPIE Vol. 199, pp 2-8, *Advances in Display Technology* (1979).

C. Reader, W. D. Flanagan, “Design Considerations for Real-Time Image Processing”, SPIE Vol. 180, *Real-Time Signal Processing II*, (1979).

C. Reader, L. Hubble, “Trends in Image Display Systems”, Proc. IEEE, Vol. 69, No. 5, pp. 606-614, (1981).

J. Adams, C. Patton, C. Reader, D. Zamora, “Hardware for Geometric Warping”, Electronic Imaging, April 1984.

Digital Image Processing Techniques (Computational Techniques), Michael P. Ekstrom (Editor), Chapter describing hardware for real-time image processing and display systems, Academic Press, 1984, pp. 289-360.

C. Reader, et al, “A New Architecture for Real-Time Image Processing,” SPIE Vol. 534, pp. 72-78, *Architectures and Algorithms for Digital Image Processing II* (1985).

B. Thacker, R. Rinaldi, C. Reader, “Multiwindow Displays of Text, Graphics and Images”, SPIE Vol. 757, pp. 104-110, *Methods of Handling and Processing Imagery* (1987).

Keynote opening speech at IEEE Conference on Image Processing, 1994.

C. Reader, “MPEG4: Coding for Content, Interactivity, and Universal Accessibility,” Optical Engineering, Vol. 35, No. 1 pp. 104-108 (1996).

C. Reader, P. Schirling, A. Puri, MPEG Tutorial paper at IEC – NCF/Infovision '97.

O. Avaro, P.A. Chou, A. Eleftheriadis, C. Herpel, C. Reader, J. Signes, “The MPEG4 Systems and Description Languages: a Way Ahead in Audio Visual Information Representation”, Signal Processing: Image Communication, Vol. 9, No. 4, pp. 385-431 (1997).

C Reader, “What MPEG-4 means to me”, Sig. Proc.: Image Comm. 9(4): 467, 1997

MPEG Video Compression Standard, Edited by Mitchell, Pennebaker, Fogg, Le Gall. Chapman & Hall, 1997. Chapter describing the process of establishing the MPEGLA patent pool, Chapter 16, *MPEG Patents*, pp.357-362.

W Gao, C Reader, F Wu, Y He, L Yu, H Lu, S Yang, T Huang, X Pan, “AVS - The Chinese Next-Generation Video Coding Standard”, Proc. NAB Conf., 1994.

C Reader, “Origins of the performance of H.264/AVC: an account of the development of H.263 and H.264 coding tools”, SPIE Vol. 5909, pp59090W-1-8 Applications of Digital Image Processing XXVIII (2005).

C Reader, “AVS Intellectual Property Rights (IPR) Policy, J. Comput. Sci. & Technol., Co. 21, No. 3 (May 2006).

Q Huang, W Gao, D Zhao, C Reader, “A Real-Time Video Deinterlacing Scheme for MPEG-2 to AVS Transcoding”, Advances in Multimedia Information Processing - PCM 2006, Vol. 4261/2006, Springer Berlin / Heidelberg.

W Gao, S Ma, C Reader, “An End-to-End Application System of AVS: AVS-IPTV in China”. PCM 2007: 50-57

C Reader, “Technical design & IPR analysis for royalty-free video codecs”. PCS 2010: 102-105

J Chen, F Xu, Y He, J D Villasenor, Y Han, Y Xu, Y Rong, C Reader, J Wen, “Efficient Video Coding Using Legacy Algorithmic Approaches”. IEEE Trans. Multimedia 14(1): 111-120 (2012)

W Gao, T Huang, C Reader, W Dou, X Chen, "IEEE Standards for Advanced Audio and Video Coding in Emerging Applications", Computer, vol.47, no. 5, pp. 81-83, May 2014.

Ma S, Huang T, Reader C, Gao W, AVS2-Making video coding smarter [Standards in a Nutshell], IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 172 - 182, 2015.

Wang R, Huang T, Park S-H, Kim J-G, Jang E S, Reader C, Gao W, The MPEG Internet Video-Coding Standard [Standards in a Nutshell], IEEE Signal Processing Magazine, Vol. 33, No. 5, pp. 164-172, Sep. 2016.

Reader C, Patent Landscape for Royalty-Free Video Coding, SPIE Vol. 9971, Sep., 2016

Patents

<u>Patent Number</u>	<u>Date Issued</u>	<u>Title</u>
6,192,073	2000	Methods and apparatus for processing video data
5,845,112	2000	Method for performing dead-zone quantization in a single processor instruction.

Academic Positions

Taught “early bird” classes at Santa Clara University in Image Processing and Digital Signal Processing

Adjunct Professor of the National Engineering Laboratory for Video Technology, in the Computer Science Department at Peking University, Beijing, China.

Professional Associations and Achievements

- ISO/IEC Certificate of appreciation for development of the MPEG4 standard.
 - Senior Member IEEE.
 - Member SPIE.
 - Member ACM.
 - Member AES.
 - Member SMPTE.
-

Contact Information

- 1654 Hillstone Ave.
Escondido
CA 92029
- 408 867 4884 Mobile-Office
- cliffreader Skype
- CliffReader WeChat
- cliff@reader.com
- www.reader.com