

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

CISCO SYSTEMS, INC.,

Petitioner,

v.

VIDEO SOLUTIONS PTE. LTD.,

Patent Owner.

IPR2024-00695

U.S. Patent No. 8,649,426

DECLARATION OF ZIXIANG XIONG, PHD

I, Zixiang Xiong, hereby declare as follows:

I. INTRODUCTION

1. I have been retained by Barnes & Thornburg LLP on behalf of Patent Owner Video Solutions Pte. Ltd. to provide my opinion concerning whether the reference(s) Petitioner relied upon in IPR2024-00695 rendered claims 1-13 (“Challenged Claims”) of U.S. Patent No. 8,649,426 (“’426 Patent,” Ex. 1001) obvious.

2. I have been retained as an independent expert consultant in this proceeding before the United States Patent and Trademark Office (PTO). Although I am being compensated at my rate of \$700.00 per hour for the time I spend on this matter, no part of my compensation depends on the outcome of this proceeding, and I have no other interest in this proceeding, any of the parties, or the subject patent.

3. A detailed description of my professional qualifications, including a listing of my specialties/expertise and professional activities and publications, is contained in my curriculum vitae (“CV”), a copy of which is provided as Ex. A to this declaration. Below is a short summary of my professional qualifications.

II. QUALIFICATIONS

4. Currently I am a full professor in and associate head of the Electrical and Computer Engineering Department at Texas A&M University. I am also the

Director of the Multimedia Laboratory at Texas A&M. I have been with Texas A&M for 25 years.

5. I have over 30 years of R&D experience on image/video processing & coding, pattern recognition & computer vision, and network communications. I have published 11 book chapters, 148 journal articles and 247 conference papers. I am also listed as an inventor on 23 patents. The details of these publications appear in my CV.

6. I received a Bachelor of Science degree in Electrical Engineering from Wuhan University, China in 1987. I was an MS student at the University of the Chinese Academy of Sciences between September 1987 and December 1989. I came to the United States in January of 1990 to further my education. I received a Master of Arts degree in Mathematics from the University of Kansas in 1991, a Master of Science degree in Electrical Engineering from the Illinois Institute of Technology in 1992, and the Doctor of Philosophy (Ph.D.) degree in Electrical Engineering in 1996 from the University of Illinois at Urbana-Champaign. In the fall of 1995, I followed my Ph.D. adviser to Princeton University, first as a visiting student (1995-1996), then as a research associate (1996-1997).

7. My academic career started at the University of Hawaii at Manoa in 1997. Two years later, I moved to Texas A&M University. I teach undergraduate courses in circuit theory, signal and systems, digital signal processing, and

probability theory. I also teach graduate courses in digital image processing & computer vision, advanced digital signal processing, random processes, information theory, pattern recognition, data compression, and network information theory and coding.

8. Because of my Ph.D. thesis work on the best wavelet-based image codec, I was fortunate to receive three prestigious research awards in my early career: an NSF Career Award in 1999, an ARO Young Investigator Award in 2000, and an ONR Young Investigator Award in 2001. These grants allowed me to fund graduate students and postdocs to work on distributed video coding. Prof. Bob Gray from Stanford University commented in 2005 that Prof. Kannan Ramchandran at UC Berkeley and I are the research leaders in distributed video coding.

9. In 2004, I was invited to write a feature article entitled Distributed source coding for sensor networks for the IEEE Signal Processing Magazine – the most cited IEEE journal according to the annual Journal Citation Report (2003 edition) published by the Institute for Scientific Information. This article subsequently received the 2006 IEEE Signal Processing Magazine best paper award. In the same year, I was elevated to an IEEE fellow for my contribution to source and channel coding. At that time, I was still an associate professor.

10. I am also co-recipient of top 10% paper awards at the 2011 and 2015 IEEE Multimedia Signal Processing (MMSP) Workshops, an IBM best student

paper award at the 2016 IEEE International Conference on Pattern Recognition (ICPR), and the best demo paper award at the 2018 IEEE International Conference on Multimedia and Expo (ICME).

11. I was a visiting professor a) in the Department of Electrical Engineering at Stanford University in Spring 2010, b) at The Methodist Hospital Research Institute in Houston, TX during 2011-2013, and (c) in the Dept. of Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia during 2016-2018.

12. I served as an associate editor for the IEEE Trans. on Circuits and Systems for Video Technology (1999-2005), the IEEE Trans. on Image Processing (2002-2005), the IEEE Trans. on Signal Processing (2002-2006), the IEEE Trans. on Systems, Man, and Cybernetics (part B) (2005-2009), the IEEE Trans. on Communications (2008-2013), and the IEEE Trans. on Multimedia (2017-2020). I also served as a guest editor for IEEE Signal Processing Magazine: Special Issue on Signal Processing for Multiterminal Communication Systems, September 2007 and IEEE Journal of Selected Topics in Signal Processing: Special Issue on MIMO-Optimized Transmission Systems for Delivering Data and Rich Content, April 2008.

13. I served as Chair of video and image processing track of IEEE ICME in 2003, Publications chair of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2007, Technical program committee co-chair,

of IEEE Information Theory Workshop (ITW) in 2007, Tutorial chair of IEEE International Symposium on Information Theory (ISIT) in 2010, Award chair of IEEE Global Communications Conference (Globecom) in 2014, and General co-chair of IEEE MMSP Workshop in 2017.

14. I was an invited Panelist on Distributed Video Coding: Trends and Challenges at the IEEE Picture Coding Symposium (PCS) in 2007. I also co-organized special sections on topics related to distributed source/video coding, distributed joint source-channel coding, and networked video at major conferences such as IEEE ICASSP, IEEE International Conference on Image Processing (ICIP), IEEE ICME, and IEEE/SPIE International Conference on Visual Communications and Image Processing (VCIP). I was also a tutorial presenter on these topics at ICASSP 2005, VCIP 2005, ICME 2007, Globecom 2007, and VCIP 2010.

15. I served on the IEEE Signal Processing Society Multimedia Signal Processing Technical Committee, the IEEE Communications Society Multimedia Communications Technical Committee, and the IEEE Circuits and Systems Society Multimedia Systems & Applications Technical Committee. I also served on the IEEE Trans. Circuits and Systems for Video Technology best paper award committee and was the best paper award chair for the IEEE Trans. Multimedia in 2017 and 2018. I was elected as a Distinguished Lecturer of the IEEE Circuits and Systems Society for 2008-2009.

16. Additional information regarding my education, technical experience and publications, etc., is included in my Curriculum Vitae attached as Ex. A to my Declaration.

17. In forming my opinion, I have relied upon my education, knowledge and experience in video encoding and technology, as well as on the materials reviewed.

III. PATENT PRINCIPLES

18. I am not a lawyer, I have not been trained in the law of patents, and I do not opine on matters of law.

19. For the purposes of my analysis, I have applied the below understanding of patent law and current legal principles related to patent validity to my analysis, as have been provided to me by counsel.

20. I have been informed by counsel that in determining whether a patent claim under post-grant review before the United States Patent and Trademark Office (PTO) is anticipated or obvious in view of the prior art, the PTO must construe the claim under the *Phillips* standard by giving the claim the meaning that the term would have to a person of ordinary skill in the art (“POSITA”) in question at the time of the invention.

21. For the purposes of this review, there are no claims at issue.

22. I have been informed by counsel 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.

23. In determining the scope and content of the prior art, I have been informed by counsel that a reference is considered appropriate 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. The fact that a reference relates to the same problem as the claimed invention supports use of the reference as prior art in an obviousness analysis.

24. To assess the differences between prior art and the claimed subject matter, I have been informed by counsel that 35 U.S.C. § 103 requires the claimed invention to be considered as a whole. This "as a whole" assessment requires considering whether 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.

25. In determining whether the subject matter as a whole would have been considered obvious at the time that the patent application was filed, by a POSITA, I have been informed by counsel of several principles regarding the combination of elements of the prior art. A combination of familiar elements according to known methods is likely to be obvious when it yields predictable results. Likewise, combinations involving 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, and choosing from a finite number of identified, predictable solutions to solve a problem are likely to be obvious. Thus, if a POSITA can implement a “predictable variation” in a prior art device, and would see the benefit from doing so, such a variation would be obvious. Also, when there is pressure to solve a problem and there are a finite number of identifiable, predictable solutions, it would be reasonable for a POSITA to pursue those options that fall within his or her technical grasp. If such a process leads to the claimed invention, then the latter is not an innovation, but more the result of ordinary skill and common sense.

26. I have been informed by counsel that the “teaching, suggestion, or motivation” test is a useful guide in establishing a rationale for combining elements of the prior art. This test poses the question as to whether there is an explicit teaching,

suggestion, or motivation in the prior art to combine prior art elements in a way that realizes the claimed invention. Though useful to the obviousness inquiry, I understand that this test should not be treated as a rigid rule. It is not necessary to seek out precise teachings; it is permissible to consider the inferences and creative steps that a POSITA (who is considered to have an ordinary level of creativity and is not an “automaton”) would employ.

27. I have been informed by counsel that when interpreting the claims of the '426 Patent I must do so based on the perspective of one of ordinary skill in the art at the relevant priority date. I have been requested by counsel to assume that the earliest claimed priority date of the '426 Patent is September 18, 2008.

IV. MATERIALS REVIEWED

28. In forming my opinions, I have reviewed and am familiar with the '426 patent and its prosecution history, the materials cited herein, and the following exhibits to the Petition.

Paper 1	The above-identified IPR Petition, along with all the ancillary documents and exhibits;
1001	The '426 Patent
1003	Declaration of Dr. Immanuel Freedman
1006	Perlman
1008	Oguz

V. LEVEL OF ORDINARY SKILL IN THE ART

29. Patent owner agrees to use Cisco's definition for a POSITA for the purposes of this IPR. *See* Pet. at 16.

VI. OVERVIEW OF THE TECHNOLOGY AND THE '426 PATENT

30. The innovations claimed in the '426 patent relate to an improvement for videoconference video data that utilizes video encoding that reduces latency and enhances resolution. The '426 patent divides the video signal into segments that are encoded independently and transmitted independently to reduce the high peak/average data bursts associated with I-frames. Ex. 1001 at Abstract, 3:29-38, 5:59-62. After the independent transmission, each segment is then decoded and the decoded/ segments are recombined. Ex. 1001 at 3:63-65.

31. The '426 patent relates to an improvement in encoding *and* transmission of video signals used in videoconferencing to reduce latency while maintaining high resolution. The '426 patent captures a stream of video data. It then separates the data into frame segments and encodes the frame segments independently and in parallel to create frame segment instances, which, after encoding, are transmitted independently over a network to a receiver such that during transmission of I-frame segment instances of one of the frame segments all other P-frame segment instances of all other frame segments are transmitted. Ex. 1001 at 7:55-8:26; 8:59-10:12.

32. The transmission limitations of claim 1 read:

- [1.8] transmitting the parallel encoded data streams independently over a network to a receiver to permit the receiver to decode each frame segment and recombine the decoded frame segments from the different encoded data streams to reproduce an image; and
- [1.9] wherein during transmission of I-frame segment instances of one of the frame segments, P-frame segment instances of all frame segments are transmitted.

Ex. 1001 at 8:19-26.

33. The transmission limitations of claim 12 read:

- [12.7]: a transceiver for transmitting the parallel encoded data streams independently over a network to a receiver to permit the receiver to decode each frame segment and recombine the decoded frame segments to reproduce an image;
- [12.8]: “wherein during transmission of I-frame segment instances of one of the frame segments, P-frame segment instances of all other frame segments are transmitted.”

Ex. 1001 at 10:5-12.

34. In this way, the '426 patent decreases latency, and reduces the overall data peak to average data burst onto the network thus maintaining high resolution by reducing packet loss. *Id.* at Figs. 3, 4; 5:50-6:2; 6:6:8-20.

35. A videoconference is “a set of interactive telecommunications technologies which allow two or more locations to interact via two-way simultaneous video and audio transmissions.” (Ex. 1001 at 1:41-44). The core technology used in a videoconference system is digital compression of audio and video streams in real time. (Ex. 1001 at 1:44-46). This compression can be understood in the context of the three major frame types found in prior art: Intra frames (I-frames), Predictive frames (P-frames), and Bi-predictive frames (B-frames)¹. *See generally* Ex. 1001 at 2:54-3:11. An I-frame (or keyframe) is a single frame of data that has all of the data needed to display that entire frame, independent of frame or frames that preceded it. (*Id.* at 2:62-65). P-frames follow I-frames and have only the data that have changed from the preceding I-frame. (*Id.* at 3:2-8). Typically, for real-time video compression only the I-frames and P-frames are considered. (*Id.* at 2:57-59). The '426 patent relates to real-time video conference methods, thus only I-frame and P-frames need to be considered.

36. Generally, it is known that the quality of the video increases when more I-frames were transmitted. Ex. 1001 at 2:66-3:1. But this quality has a cost.

Because the I-frames include data for an entire video frame, encoding I-frames required significantly more time, thus increasing latency. Also transmitting the I-frame would increase the network traffic load and create high peak/average data bursts on the network, leading to choppy video and/or packet loss due to delays in receiving the I-frame data, ultimately degrading video quality. Ex. 1001 at 3:29-36. Such bursts are depicted in Figure 2 below by the “No Segments” spikes of 60 kB.

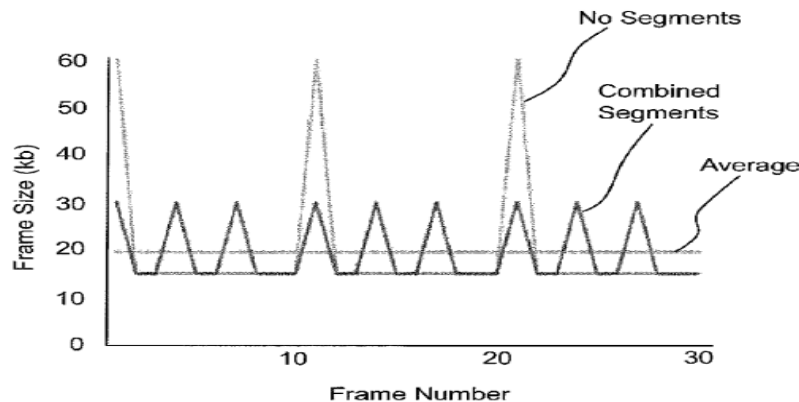


Figure 2

PRIOR ART

37. Because the I-frames have a lot of data, encoding I-frames took significant time thus increasing latency. As videoconferencing and telepresence was becoming more popular, the need to transmit large amounts of video data consistently and in real-time was becoming a larger problem. (Ex. 100 at 2:33-36).

38. “Telepresence” is a “set of technologies which allow a person to feel as if they were present, to give the appearance that they were present, or to have an effect, as a location other than their true location.” (Ex. 1001 at 2:20-23).

39. Many prior art encoding techniques were developed to overcome these obstacles including encoding techniques, such as MPEG and H.264, to reduce the amount of video data being transmitted. (Ex. 1001 at 2:36-38).

40. MPEG, for example, achieved a high compression rate by representing only the changes from one frame to another, instead of the entire frame, and then encoding the video information using Discrete Cosine Transfer (DCT). *See* Ex. 1001 at 2:40-44. At the time of the invention, DCT was a widely used technique in data compression and uses lossy compression techniques to reduce the data size or transmission, achieving higher compression ratios. Also at the time of the invention, similarly, H.264 or Advanced Video Coding was a known standard video compression format. It relies on similarities between frames and uses those similarities to reduce the amount of data needing to be stored and transmitted. H.264 also uses lossy compression, where some data is removed, but the effect is generally not perceived by the human eye. *See* Ex. 1001 at 51-54.

41. These techniques proved suitable for one-way (*i.e.*, single direction) applications such as broadcast, video on demand, and DVD video storage applications. (*See* Ex. 1001 at 5:31-35). In these applications, encoding is typically

performed ahead of time such that processing equipment/time does not significantly contribute to latency. *Id.* at 5:35-37. Also, in these applications, the video data is buffered before playback so as to alleviate problems relating to high peak data bursts. (See Ex. 1001 at 5:31-49).

42. The prior art encoding techniques available at the time of the invention could achieve compression ratios of 200:1 or even 500:1, but had some drawbacks. Particularly, they were very expensive computationally (requiring greater computer processing power and long wait times to render a video) such that PC-based videoconferencing system did not have enough processing capabilities to handle video encoding operations in real time. (Ex. 1001 at 3:12-23). Also, as noted above, the prior art encoding techniques resulted in “bursty” data, resulting in choppy video and/or packet loss due to delays in receiving the I-frame data. (Ex. 1001 at 3:29-36).

43. In my opinion, the inventors of the '426 patent found a solution to the problems described above. They found that that by dividing the video signal into multiple, separate streams, they could create video segments identical from one video frame to the next frame. To reduce latency, the separate streams could then be encoded in parallel. The inventors also realized that by encoding and transmitting independently the parallel encoded streams over the network, and by ensuring that during the transmission of an I-frame segment from one of the streams, all of the P-

frame segments were also transmitted, all of the segments for each frame would be transmitted before the next frame resulting in reduced network congestion and decreased latency.

44. Thus, the invention's focus on videoconferencing allows the use of an I-frame segment instance and one or more P-frame segment instances to exploit the fact that large areas of each frame will not change frame to frame. Ex. 1001 at 6:66-7:10 ("the video encoder does not need to determine this fact in an analytical manner for each and every frame, at a rate of 24 to 60 frames per second. Rather, it is necessary only for the video encoder to process the image to determine the regions of movement."). "Optionally, the rate at which I-frames are sent out is reduced relative to that of general purpose video encoding. This is because, unlike movies and broadcast television, there is a low probability in videoconferencing that the video signal will include a scene change, panning motion, or other type of change that requires a new I-frame to be sent." Ex. 1001 at 7:3-11.

45. By transmitting the parallel encoded data streams independently and by transmitting P-frame segment instances of all other frame segments during the transmission of I-frame segment instances for one of the frame segments, the patent supports independent reconstruction of the segments at the distant end. Ex. 1001 at Abstract, 3:29-38; 5:59-62; 7:55-8:27; 8:59-10:12. After the independent

transmission, each segment is then decoded and the decoded segments are recombined. Ex. 1001 at 3:63-65.

46. Referring to Figure 3 reproduced below, raw high-resolution video signal is sent to a divider, which divides each frame of the video signal into a plurality of segments. (Ex. 1001 at 5:50-55). Each segment is then encoded independently using, for example, a separate encoder or other type of parallel processor, which reduces latency. (*Id.* at 5:66-6:1).

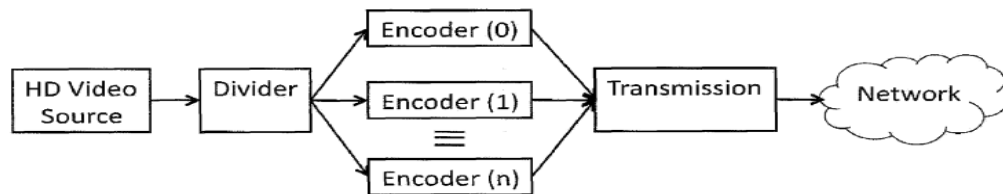


Figure 3

Ex. 1001 at Fig. 3.

47. Also, because each segment is encoded separately, the I-frame occurrence of each segment is preferably offset from the other segments, such that each segment's I-frame occurs when at least another segment produces a P-frame, thereby reducing the overall peak to average data burst on the network and thus reducing packet loss. (Ex. 1001 at 6:7-12). The effect of separate encoding and independent transmission is shown in Figure 4 below where the I-frames for

segments (0), (1), and (2) are shown offset from one another. (Ex. 1001 at 6:12-14); Figs. 3 and 4.

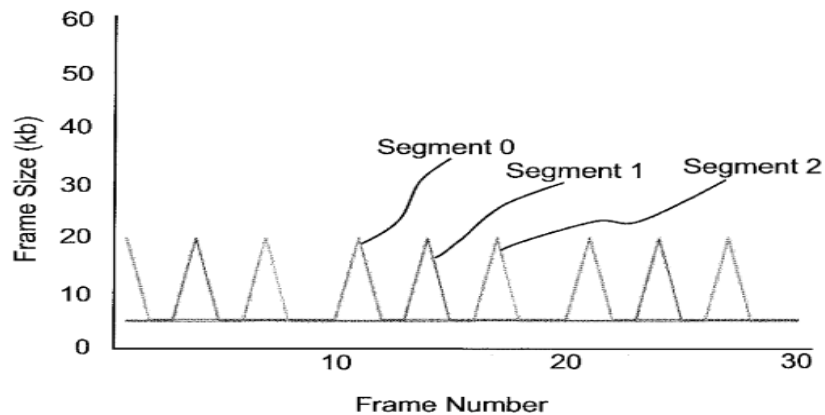


Figure 4

48. Referring to Figure 4, each signal has a P-frame size of ~5kB and an I-frame size of ~20kB. For the three segments (0, 1, and 2), 10kB of P-frame data is sent for each frame and 20kB of I-frame data is sent for the same frame resulting in 30kB of frame data, which is shown in the Figure 2 “Combined Segments” graph.

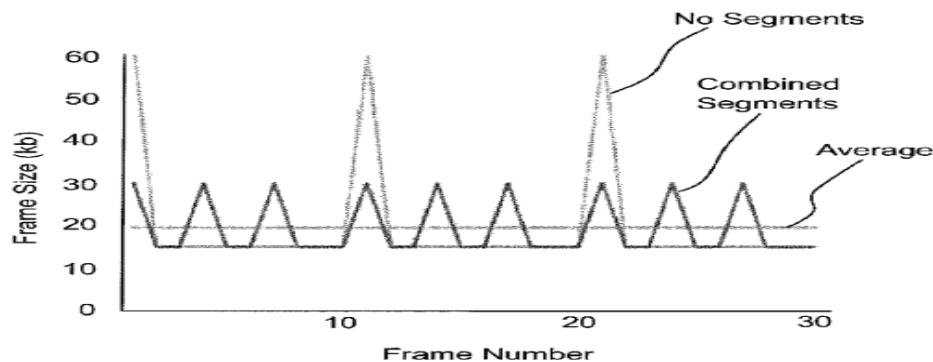


Figure 2
PRIOR ART

Ex. 1001 at Fig. 2.

49. The invention of the '426 patent reduces the burst data bit rate (compare prior art “No Segments” peaks with the invention’s “Combined Segments” peaks on the Figure 2 graphs) while maintaining a similar average data rate and a similar video transmission encoding quality. (Ex. 1001 at 6:20-23).

VII. SUMMARY OF THE ASSERTED PRIOR ART

a. Perlman (Ex. 1006)

50. Perlman relates to methods of broadcasting a live game tournament in the form of streaming interactive video from a hosting service to a plurality of viewers. Ex. 1006; Abstract. Perlman identifies many obstacles in the context of on-line gaming, including that the game logic may take shorter or longer than one frame time to compute the graphics displayed for the next frame depending upon the complexity of the scene (*id.* at ¶ [0036]); and making live video feeds available to

the television channel's video distribution and processing equipment in real-time, particularly in a tournament situation. (*Id.* at ¶ [0040]).

51. Perlman describes multiple approaches to addressing these and other obstacles. One approach is to host the gaming program code and data on a server and deliver the data to client machines on demand as compressed video and audio streamed over a digital broadband network. (*Id.* at ¶ [0057]). This approach, however, has a major drawback, *i.e.*, latency (the time it takes for a signal to travel to and from the game server). (*Id.*). Latency is particularly a problem in so-called fast-action or “twitch” video games. *Id.* Perlman identifies multiple sources of latency, including: round-trip latency between the user's home and the general Internet, routing delays, data rate capacity, and the time necessary to compress and decompress the video. (*Id.* at ¶¶ [0059, 0060, 0061]).

52. Perlman focuses on improving latency in online video games. Perlman distinguishes between compression from so-called “tame” video streams that do not change very much (*e.g.*, as would “be common in video teleconferencing where the cameras are in a fixed position and have little motion, and the objects, in the scene, *e.g.*, seated people talking, show very little motion”), and twitch video games. (*Id.* at ¶¶ [0163, 0164]; *compare* Figs. 6(b) and 6(c)). Perlman teaches that its encoding scheme, unlike prior art Constant Bit Rate (CBR) encoding, is not necessary for

videoconferencing applications, but rather should be used for rapidly changing scenes such as those in twitch video games. Ex. 1006 at [0167].

53. Perlman discloses a variety of solutions to reduce latency in the video game context. While Perlman does discuss reducing so-called I-frame peaks (*see* Ex. 1006 at Fig. 6b; Pet. at 23-24) by dividing frames into tiles consisting of I-frame tiles and P-frame tiles (*see* Fig. 7a), Perlman does not disclose or even consider the transmission aspects of the '426 patent's invention. None of the figures and text from Perlman relied on by Petitioner as allegedly teaching limitations [1.8] and [1.9] (and [12.7] and [12.8]) relate to transmission. *Compare* Petition, at 47-55 (discussing Figs. 7a, 10b, and 12) with Perlman [0083] – Figs. 7a-b illustrate video *compression* techniques, [0086] – Figs. 10a-b illustrate efficiently *pack[ing]* image tiles within packets, [0088] – Fig. 12 – multi-core processing units for *compression*. Perlman, therefore, at most discloses the concept of dividing a frame into segments or tiles, but that is only a portion of features recited in the claims of the '426 patent. In my opinion, Perlman does not teach, and Petitioner fails to demonstrate, that Perlman discloses and/or that a POSITA would find obvious, the '426 patent's *transmission* limitations as discussed in detail below.

b. Oguz (Ex. 1008)

54. Oguz relates to techniques for content adaptive video frame slicing and non-uniform access unit coding for improved coding efficiency.” Ex. 1008, Abstract.

It is my understanding that Petitioner does not argue, and Oguz does not teach, the '426 patent's transmission limitations.

VIII. CLAIM CONSTRUCTION

55. I understand that claim terms, in the absence of an express definition or disclaimer, are given their ordinary and customary meaning, as would have been understood by a POSITA at the time of the alleged invention.

56. In considering the meaning of the claims, however, I understand that one must consider the language of the claims in light of “the remainder of the specification, the prosecution history, and extrinsic evidence concerning relevant scientific principles, the meaning of technical terms, and the state of the art.” 37 C.F.R. § 42.100(b).

57. I understand that Petitioner did not propose construction for any term used in the claims of the '426 patent.

IX. THE CHALLENGED CLAIMS

58. Petitioner challenges claims 1-13. *See* Paper 1 at 25.

59. Independent Claim 1 of the '426 Patent recites as follows (delineations 1.0, 1.1, 1.2, etc. are added for convenience).

1.0	A method of encoding in real time a video signal that is associated with a videoconference or telepresence session, the method comprising:
1.1	capturing a stream of video data comprising a series of sequential video frames;

1.2	separating the video frames into a plurality of frame segments, wherein each frame segment represents a part of a complete video frame;
1.3	encoding each frame segment independently and in parallel to generate parallel encoded data streams, one for each of the plurality of frame segments,
1.4	each encoded data stream containing a sequence of frame segment instances each corresponding to a particular frame segment in the video frames, and
1.5	each encoded data stream containing I-frame segment instances that carry complete information about a particular frame segment and P-frame segment instances that carry partial information about the particular frame segment, the partial information representing information that has changed relative to a previous frame segment instance in that encoded data stream, and
1.6	wherein the encoding of the frame segments for each frame is performed such that when an instance of one of the frame segments within a particular frame is an I-frame segment instance, at least one of the other frame segment instances in that particular frame is a P-frame segment instance,
1.7	whereby in the respective parallel encoded data streams in any period of time when an I-frame segment instance is present in one of the parallel encoded data streams, a P-frame segment instance is present in at least one of the other parallel encoded data streams;
1.8	transmitting the parallel encoded data streams independently over a network to a receiver to permit the receiver to decode each frame segment and recombine the decoded frame segments from the different encoded data streams to reproduce an image;
1.9	wherein during transmission of I-frame segment instances of one of the frame segments, P-frame segment instances of all other frame segments are transmitted.

60. Independent Claim 12 of the '426 Patent recites as follows (delineations 12.0, 12.1, 12.2, etc. are added for convenience).

12.0	A video encoding system for encoding in real time a video signal that is associated with one of a videoconference and a telepresence session, the system comprising:
12.1	a port for receiving a stream of video data comprising a series of sequential video frames;
12.2	a divider for dividing the stream of video data into a series of substreams of data corresponding to frame segments, wherein each frame segment represents a part of a complete video frame;
12.3	plurality of video encoders arranged in parallel for independently encoding the respective frame segments to generate parallel encoded data streams, one for each of the plurality of frame segments,
12.4	each encoded data stream containing I-frame segment instances that carry complete information about a particular frame segment and P-frame segment instances that carry partial information about a particular frame segment, the partial information representing information that has changed relative to a previous frame segment instance in that encoded data stream, and
12.5	the video encoders being arranged to encode each frame segment such that when an instance of one of the frame segments within a particular frame is an I-frame segment instance, at least one of the other frame segment instances in that particular frame is a P-frame segment instance,
12.6	whereby in the respective parallel encoded data streams in any period of time when an I-frame segment instance is present in one of the parallel encoded data streams, a P-frame segment instance is present in at least one of the other parallel encoded data streams; and
12.7	a transceiver for transmitting the parallel encoded data streams independently over a network to a receiver to permit the receiver to decode each frame segment and recombine the decoded frame segments to reproduce an image;
12.8	wherein during transmission of I-frame segment instances of one of the frame segments, P-frame segment instances of all other frame segments are transmitted.

X. PETITIONER MISINTERPRETED AND/OR IGNORED THE TEACHINGS IN PERLMAN

61. It is my opinion that Dr. Freedman and Petitioner misinterpreted and/or ignored the teachings in Perlman. Perlman, in my opinion, does not disclose or suggest to a POSITA claim elements [1.8], [1.9], [12.7], or [12.8]. These elements relate to “*transmitting* the parallel encoded data streams *independently* over a network ... [1.8]” and “wherein during *transmission* of I-frame segment instances of one of the frame segments, P-frame segment instances of all frame segments are *transmitted* [1.9].” Ex. 1001 at 8:19-27. Perlman does not teach either limitation. The parts of Perlman relied on by Petitioner do not even relate to transmission. Thus, in my opinion, Petitioner fails to establish that independent claim 1 is unpatentable over Perlman as applied in Ground 1, and because Ground 2 relies on the analysis of independent claim 1, Perlman does not teach that Ground as well.

62. My review of the Petition reveals that Petitioner makes three arguments that Perlman discloses element [1.8]. *See* Pet. at 47-52. The Petition makes these same arguments for element [12.7].

63. Petitioner cites to paragraphs [0214] and [0117] and argues that these paragraphs teach that that Perlman’s compressed tiles are transmitted to a destination. Pet. at 48. These paragraphs, however, do not teach or otherwise indicate “transmitting the parallel encoded data streams *independently*...” as required by the claims. The figures and text referred to in the paragraphs (Fig. 12

for paragraph [0214]² and Figs. 4a, 2a, and 2b for paragraph [0117] do not provide any further teaching. Figure 12, for example, “illustrates one embodiment which uses multi-core processing units for compression.” Ex. 1006, at [0088]. Figures 2a and 2b are high level system architecture illustrations. *Id.* at [0071]. And figure 4a illustrates a hosting service and a client. *Id.* at [0073]. In short, none of the Petition’s cited evidence teach claim elements [1.8] or [12.7].

64. Petitioner advances a second argument regarding the “independently” transmitted requirement of the claims, *i.e.*, that “Perlman’s streams are ‘*independently*’ transmitted because they have to be recombined on the receiving end.” Pet. at 48-49. Citing Perlman paragraph [0170] discussing decompression of I frames and P frames, Petitioner argues that multiple instances of a video decompressor is known in the art. Petitioner also argues (without citation) that “the compressed tiles maintain their independence through transmission.” Pet. at 48. That the tiles are decompressed *after they are transmitted* says nothing about whether or not they were transmitted *independently* as required by the claim.

65. A simple example illustrates the fault in Petitioner’s logic. Suppose four people (each representative of a tile of a frame) were to travel (be transmitted) from Dallas (where encoding/compression occurred) to Washington, D.C., where

² Figure 12 is referred to in paragraph [0213] and paragraph [0214] refers to “techniques previously described.” Ex. 1006, at [0214].

decompression occurs. That the four people arrived in Washington, D.C. says nothing about how they traveled to Washington. They could have traveled in separate cars (packets) on separate roads, or in separate cars on the same road each behind the other. They could have traveled in pairs in two cars. Three could have traveled together and the fourth separately, or all four could have traveled in one car. They could have arrived simultaneously, or one could have arrived before the others and waited (*e.g.*, been buffered/held) while the others arrived. In short, that Perlman decompresses each tile separately and reconstructs the video on the receiving end says nothing about the claims' required *independent transmission* limitation, which is a *separate step from decompression*.

66. Petitioner advances another alternative argument for the independent transmission limitation—"Perlman's tiles are transmitted '*independently*' because different tiles—belonging to different data streams—are transmitted in separate packets." Pet. at 49-50. Petitioner cites to Figure 10*b* in support of this argument, arguing that because at least some of the tiles (*e.g.*, T7 and T2) are transmitted in separate packets "at least some of the encoded streams are transmitted '*independently*' from each other." Pet. at 50.

67. This argument ignores Perlman's teaching. Figure 10*b* and the tile packing logic described in [0198] demonstrate *dependent transmission*, not "independent transmission" as required by the claims. Perlman explains that Fig.

10b illustrates tile packing logic for intelligently packing tiles within packets to reduce the effect of packet loss. Ex. 1006 at [0198]; *see also* [0199]. Tile packing logic 1010 aligns tiles with packet boundaries (*e.g.*, tiles T1, T3, T4, T7, and T2 are aligned with packet boundaries). *Id.* Then, based on the size of the tiles, T1 and T6 are combined in packet 1001, T3 and T5 are combined in packet 1002, and so on. *Id.* This scheme is the opposite of independent transmission because certain tiles are transmitted together (dependent) in packets and the purpose of the tile packing logic is to best fit the tiles into packets based on assigned rules (*i.e.*, tile size and packet size).

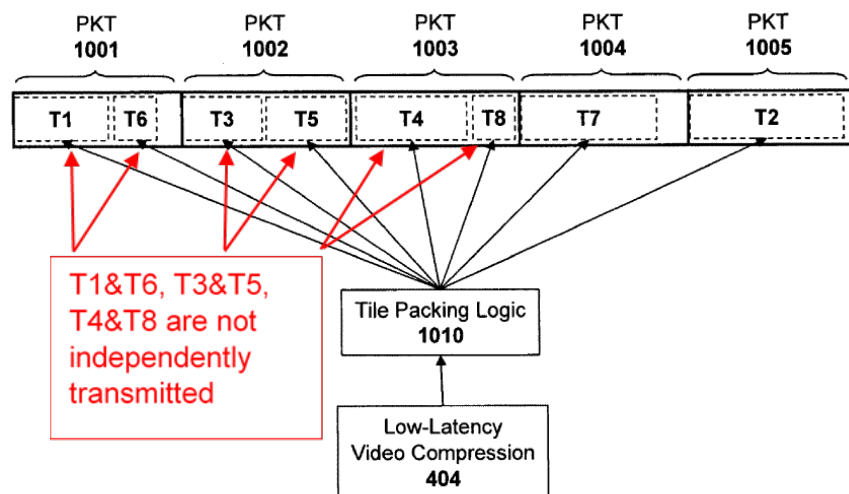


Fig. 10b

68. The annotations to Perlman Fig. 10b illustrate the effect of these rules, *i.e.*, that certain tiles are packaged together for later transmission. A POSITA

reading Perlman would, therefore, conclude that Perlman disfavors from independent transmission.

69. Perlman characterizes as “inefficient[]” packing, *i.e.*, where a tile crosses a packet boundary (Fig. 10a), because the loss of a packet can result in a loss of multiple frames. Ex. 1006 at [0197]. Perlman, therefore, undermines the very argument Petitioner makes in support of independent transmission—that a POSITA would have found it obvious for each tile to be transmitted as a separate packet. Dr. Freedman’s declaration provides no non-conclusory reasoning why a POSITA would have found independent transmission obvious. In paragraph 83 of his declaration, Dr. Freedman simply states that because some tiles are transmitted in separate packets at least some of the encoded streams are transmitted independently, but the claim does not say some streams are transmitted independently. As discussed, Perlman teaches that packing logic should be employed to combine the output of streams to fill the packets, which is not independent transmission. Based on this teaching, I disagree that a POSITA reading Perlman would have found it obvious for each tile to be transmitted in a separate packet as Dr. Freedman suggests.

70. Finally, Petitioner relies on Perlman [0170] to argue the last portion of claim element [1.8] is met and analogizes Perlman to Figure 3 of the ’426 patent. Pet. at 51-52. But as discussed above, that the streams in Perlman are ultimately decompressed fails to provide any information regarding their transmission prior to

the decompression step. Also, as discussed above, Perlman teaches tile packing logic, which occurs prior to transmission, and, assuming transmission occurs thereafter, would render such transmission dependent because the outputs of multiple streams would be transmitted together.

71. In my opinion, Perlman also does not disclose claim elements [1.9] and [12.8]. The Petition relies on paragraph [0169] and Figs. 7a and 12 to argue that Perlman teaches this limitation. Pet. at 52-54.³ None of these citations relate to *transmission*, however. Each relates to encoding/compression, which is a different step in the process.

72. Paragraph [0169], for example, teaches that “low-latency compression logic 404 divides up the video frame into a number of tiles....” These tiles are represented in Fig. 7a, which illustrates “example video compression techniques.” Ex. 1006 at [0083]. Figure 12 “illustrates one embodiment which uses multi-core processing units for *compression*.” *Id.* at [0088]; *see also* [0212, 0213] (discussing compression). The Petition’s *transmission-related* annotations are not present in either figures 7a or 12, and are not described in the text of Perlman. The cited portions Dr. Freedman’s declaration (¶¶ 88-90) also do not provide any additional insight. Dr. Freedman simply repeats the Petition and reproduces paragraph [0169] of Perlman. Perlman, therefore, does not disclose either of limitations [1.9] and

³ The Petition makes the same arguments for claim limitation [12.8]. Pet. at 76.

[12.8]. To conclude, the Petition offers no evidence that Perlman teaches this step or that a POSITA would find these limitations obvious. As discussed above, Perlman seemingly teaches away from these limitations by imposing tile packing logic rules based on tile size and available packet space.

XI. JURAT

73. In signing this Declaration, I understand that the Declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I acknowledge that I may be subject to cross-examination in this case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination in person or remotely within the United States during the time allotted for cross-examination.

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

Executed on this _27th_ day of ____June____ 2024

A handwritten signature in cursive script, appearing to read "Zixiang Xiong".

Zixiang Xiong