UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC., AMAZON.COM SERVICES, LLC,
Petitioner

v.

NOKIA TECHNOLOGIES OY,
Patent Owner

Case: IPR2024-00627
U.S. Patent No. 11,805,267

DECLARATION OF DR. IAIN RICHARDSON

TABLE OF CONTENTS

I. Introduction

1. I have been retained on behalf of Nokia Technologies Oy and asked to provide opinions in respect of two Inter Partes Review (IPR) proceedings, IPR2024-00626 and IPR2024-00627, both of which relate to U.S. Patent No. 11,805,267 (the "'267 Patent"). I understand that the Board determined that IPR should be instituted in IPR2024-00626 and IPR2024-00627.

2. This Declaration summarises the independent opinions I have formed to date based on the material currently available to me. I reserve the right to modify my opinions, if necessary, based on further review and analysis of information I receive after the filing of this Declaration, including in response to any positions taken by Petitioners or their experts that I have not yet been provided with.

3. My findings, as explained below, are based on my study, experience, knowledge, and background in the fields discussed below. I am informed by my education and my extensive experience in the fields of video compression and multimedia coding and communications. I have also relied on my review and analysis of the prior art, information provided to me and information I have independently reviewed.

4. I am submitting the same declaration for both IPR2024-00626 and IPR2024-00627, as these petitions raise the same grounds for encoding (-00626) and

decoding (-00627) claims in the '267 Patent. Where applicable, my cites herein are with respect to the exhibits and page numbers reflected in IPR2024-00626.

II. Educational Background and Qualifications

5. Details of my professional qualifications and background are summarized below and are set out in detail in my *Curriculum Vitae*, attached to this declaration.

6. I am Director and co-founder of Vcodex BV.

7. I have over 30 years of experience related to video and image compression and coding. I am qualified by education and experience to testify as an expert with respect to subject matter in the fields of video, image, multimedia and general-purpose data encoding, decoding, compression, decompression, and transmission.

8. I received a Master of Engineering (M.Eng.) degree in Electronic and Electrical Engineering from Heriot-Watt University in Edinburgh, Scotland in 1990. I received a Doctor of Philosophy (Ph.D.) degree in Video Compression from Robert Gordon University in Aberdeen, Scotland in 1999. I worked as a Digital Signal Processing ("DSP") Hardware Designer with GEC Avionics Ltd. from 1990 to 1993.

9. In 1993, I assumed a post as a Lecturer, then Reader and eventually a Full Professor in the field of image and video compression in the School of

Engineering at Robert Gordon University. In 2009, I was awarded the position of Honorary Professor with Robert Gordon University.

10. At Robert Gordon University, I founded and ran an image communication technology research laboratory. I carried out original research in the field of data, image, and video compression, initiated and managed research projects, and supervised research students.

11. I am the author of five books and over 70 journal and conference papers on video and image coding and communication, including two widely cited books on the H.264 / MPEG-4 industry standards for video compression. In my 2002 book, *Video Codec Design*, I discussed in detail the efficient implementation of image and video compression on software and hardware platforms.

12. In 1995, during my employment at Robert Gordon University, I cofounded and became Technical Director, equivalent of Chief Technology Officer, at 4i2i Communications Limited. I held this position until 2000. In my role at 4i2i, I was responsible for and engaged in the design of software and hardware modules for compression and communication systems. The designs included video processing algorithms in software and hardware for implementation on processing devices and the design of compression algorithms such as block-based transforms, run-length coding, and variable-length coding in software and hardware. A modified version of

one of my video compression hardware designs was successfully used by NASA to stream video from the International Space Station.

13. In 2007, I founded Vcodex Limited and currently serve as the CEO, through which I provide technical and strategic advice to technology companies particularly relating to video, image, and data compression for streaming.

14. In 2009, I founded Onecodec Limited and served as CEO, leading the company's development of innovative video, image, and multimedia data compression and storage software and systems. Onecodec's technology and business activities were merged with Vcodex Limited in 2015.

15. I have previously prepared expert reports and testified in a number of cases concerning video and image compression, graphics compression, general-purpose data compression, and multimedia streaming.

16. I am a named inventor on 11 patents or patent applications, including U.S. and non-US patents and applications.

17. Because of my background, training, and experience, I am qualified as an expert in the technical areas relevant to the '267 patent.

18. My company, Vcodex BV, is being compensated for my independent analysis as an expert with respect to this inter partes review (IPR) proceeding, but neither my compensation nor Vcodex BV's compensation is contingent in any way on the content of my analysis or the outcome of this proceeding. I have no financial

interest in Petitioners (Amazon.com, Inc., Amazon.com Services LLC) or Patent Owner (Nokia Technologies Oy) and I have no financial interest in the challenged patent.

III. Materials Reviewed

19. The opinions set forth in this Declaration are based on my professional judgment, education, experience, and knowledge regarding the video coding technologies that are discussed in the ’267 Patent. In addition, I have based my opinions on my review of documents, including those listed below:

Exhibit	Title
2001	Declaration of Dr. Iain Richardson
2002	CV of Dr. Iain Richardson
2003	File History of App. Ser. No. 15/250,124 (“’124 Application”)
2004	File History of App. Ser. No. 15/490,469 (“’469 Application”)
2005	File History of App. Ser. No. 15/876,495 (“’495 Application”)
2006	File History of App. Ser. No. 16/729,974 (“’974 Application”)
2007	Pat. App. Pub. No. US 2009/0257503 to Yan Ye, Peisong Chen, and Marta Karczewicz (“Ye ’503”)
2008	Pat. App. Pub. No. US 2010/0086027 to Rahul Panchal and Marta Karczewicz (“Panchal”)
2009	Pat. App. Pub. No. US 2013/0142262 to Yan Ye and Alexandros Tourapis (“Ye ’262”)
2010	VCEG-AI33 by Yan Ye, Peisong Chen, and Marta Karczewicz
2011	U.S. Pat. No. 9,161,057 to Marta Karczewicz, Peisong Chen, and

	Yan Ye (“Karczewicz ’057”)
2012	Excerpts from “The H.264 Advanced Video Compression Standard”, Richardson, 2010 (Cover pages, book pages 28-38, 69, 106, 108, 223-229)
2016	Deposition Transcript of Immanuel Freedman, dated November 26, 2024
2017	Marta Karczewicz 50 Leading Tech Voices, available at https://www.epo.org/en/about-us/50-epc-anniversary/50-leading-tech-voices/marta-karczewicz
2018	Richardson, The H.264 Advanced Video Compression Standard (page 283)

IV. Legal Standards

20. I am not an attorney or a legal expert, and I do not opine on matters of law. All the legal standards, which are listed in this section and which I applied in reaching the conclusions expressed throughout this Declaration, were provided to me and explained to me by counsel. In preparing this Declaration, I have endeavored to faithfully apply these legal standards. I do not intend to offer opinions on the law. Below, I describe the legal standards that I applied in my analysis of the claims of the ’267 Patent.

21. I have been informed by counsel that an issued patent is presumed valid and that the challenger of a patent’s validity bears the burden of proving invalidity of the patent.

A. Patent Claims

22. I have been informed by counsel that a patent's claims describe the invention made by the inventors and describe what the patent owner owns and what the owner may prevent others from doing.

23. I have been informed by counsel that claims may describe products, such as machines or chemical compounds, or processes for making or using a product. Claims are usually divided into parts or steps called limitations.

24. I have been informed by counsel that there are two types of claims: independent claims and dependent claims. An independent claim sets forth all the requirements that must be met in order to be covered by that claim. Thus, it is not necessary to look at any other claim to determine what an independent claim covers. A dependent claim does not itself recite all of the requirements of the claim but refers to another claim for some of its requirements. In this way, a dependent claim depends on another claim. The law considers a dependent claim to incorporate all of the requirements of the claims to which it refers.

25. I have been informed by counsel that that a dependent claim then adds its own additional requirements. To determine what a dependent claim covers, it is necessary to look at both the dependent claim and any other claims to which it refers.

26. I have been informed by counsel that in a case where a dependent claim incorporates the limitations of an independent claim upon which it depends, if a

reference fails to disclose any of the limitations of the independent claim, the claims which depend on that independent claim cannot be made invalid by such a reference. As a result, any of the references at issue in this proceeding that fail to disclose all of the limitations of an independent claim necessarily also fail to disclose all of the limitations of claims which depend on that independent claim.

B. Burden of Proof for Invalidity

27. I have been informed by counsel and I understand that the party challenging the validity of an issued U.S. patent in an IPR bears the burden of proving invalidity through facts supported by a preponderance of the evidence. I

28. have been informed that a "preponderance of the evidence" means evidence sufficient to prove that a fact is more likely true than it is not.

C. Claim Construction

29. I have been informed by counsel and I understand that the first step in determining validity is to properly construe the claims. I have further been informed by counsel that, in this proceeding, the Board must determine the scope of the claims by using the same claim construction standard that United States District Courts use. Under this standard, the words of a claim are given their "ordinary" and "customary" meaning as would have been understood by a person of ordinary skill in the art at the time of the invention. I have been informed that such a person looks to the claims,

specification, and file history in understanding the meaning of the claim terms. I have applied this understanding in this declaration.

D. Person of Ordinary Skill in the Art

30. I have been informed by counsel and I understand that a person having ordinary skill in the art may be defined by factors such as, when considered at the time of the alleged invention: (a) the levels of education and experience of the inventor and other persons actively involved in the field; (b) the types of problems encountered in the field; (c) prior art solutions to those problems; (d) rapidity with which innovations are made; and (e) the sophistication of the technology.

E. Anticipation

31. I have been informed by counsel and I understand that a patent claim can be found invalid as anticipated if each and every element of the claim, as properly construed, is found either explicitly or inherently in a single prior art reference. If the single prior art reference does not disclose explicitly or inherently any one claim limitation, then that claim is not anticipated. I have further been informed that the single prior art reference cannot merely disclose each element; rather, it must disclose all of the elements as arranged in the claim. I have been informed that for a prior art reference to "inherently" disclose something, the reference must necessarily function in accordance with, or include the claimed limitations, to anticipate. The missing descriptive matter cannot merely probably or

possibly include the claimed limitations. I have further been informed by counsel that, with respect to an anticipation analysis, it is improper to combine separate, distinct embodiments of a single prior art reference.

32. I have been informed that for a reference to be anticipatory, the reference must contain an enabling disclosure. I have been informed that merely naming or describing the subject matter is insufficient if it cannot be produced without undue experimentation.

F. Obviousness

33. I have been informed by counsel that a patent claim can be found invalid as obvious if the differences between the claim and the prior art references would have been obvious to a person of ordinary skill in the art at the time of the invention, taking into account (1) the scope and content of the prior art, (2) the differences between the prior art and the claims, (3) the level of ordinary skill in the art, and (4) any objective indicia of non-obviousness. I have further been informed by counsel that a claim can be found to be obvious in light of either a single prior art reference or multiple prior art references. I have been informed that references qualify as prior art for an obviousness determination only when they are analogous to the claimed invention. Two separate tests define the scope of analogous prior art: (1) whether the art is from the same field of endeavor, regardless of the problem addressed; and (2) if the reference is not within the field of the inventor's endeavor,

whether the reference still is reasonably pertinent to the particular problem with which the inventor is involved.

34. I have been informed that to be obvious in light of a single prior art reference or multiple prior art references, there must be a reason either to modify the single prior art reference or to combine the two or more prior art references to achieve the claimed invention. I have been informed by counsel that this reason may come from a teaching, suggestion, or motivation to combine which may come from the reference or references themselves, the knowledge or "common sense" of one skilled in the art, or from the nature of the problem to be solved, and may be explicit or implicit from the prior art as a whole.

35. I have been informed by counsel that the combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.

36. I have further been informed by counsel that there is no suggestion to combine, however, if a reference teaches away from its combination with another source. A reference may be said to teach away when a person of ordinary skill, upon reading the reference, would be discouraged from following the path set out in the reference, or would be led in a direction divergent from the path that was taken by the applicant, or if it suggests that the line of development flowing from the reference's disclosure is unlikely to be productive of the result sought by the

applicant. If, when combined, the references would produce a seemingly inoperative device, then they teach away from their combination. I have been informed that it is improper to combine references where the references teach away from their combination.

37. I have been informed by counsel that the degree of teaching away will depend on the particular facts; in general, a reference will teach away if it suggests that the line of development flowing from the reference's disclosure is unlikely to be productive of the result sought by the applicant. I have been informed that a reference teaches away, for example, if (1) the combination would produce a seemingly inoperative device, or (2) the references leave the impression that the product would not have the property sought by the applicant.

38. I have been informed by counsel that if the teachings of a prior art reference would lead one skilled in the art to make a modification that would render another prior art device inoperable, then such a modification would generally not be obvious. I have been informed that if a proposed modification would render the prior art device unsatisfactory for its intended purpose, then the proposed modification would generally not be obvious.

39. I have been informed by counsel that it is improper to use hindsight in making the obviousness determination.

40. I have been informed by counsel that some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness must be made. I have been informed that, for example, there may be no rational underpinning to support the legal conclusion of obviousness where systems are already complete and the party challenging validity seeks to mix and match components, which each exist in the separate systems, from these complete systems.

V. Person of Ordinary Skill in the Art

41. Based on my review of the ’267 Patent, the types of problems encountered in the art at and prior to the time of the alleged invention, prior solutions to those problems, the pace with which innovations were made during the relevant time period, the sophistication of the technology, and the educational level of active professionals in the field, I believe that a person of ordinary skill in the art, at the time of the invention, would have had at least a Bachelor’s degree in electrical engineering, computer science, or computer engineering; familiarity with the relevant video coding standards as of this time; and at least two years of work experience in research, design, development, and/or testing of video coding hardware and/or software. More education could have substituted for less work experience and vice versa.

42. All my opinions in this Declaration are from the perspective of a person of ordinary skill in the art, as I have defined it above and during the relevant time

frame (e.g., 2011). During that time frame, I possessed at least the qualifications of a POSITA, as defined above. I also note that Dr. Freedman is familiar my work in video coding and described my 2010 book[1] as being heavily cited. Ex-2016 at 108:19-110:1.

VI. Background of the Relevant Technology

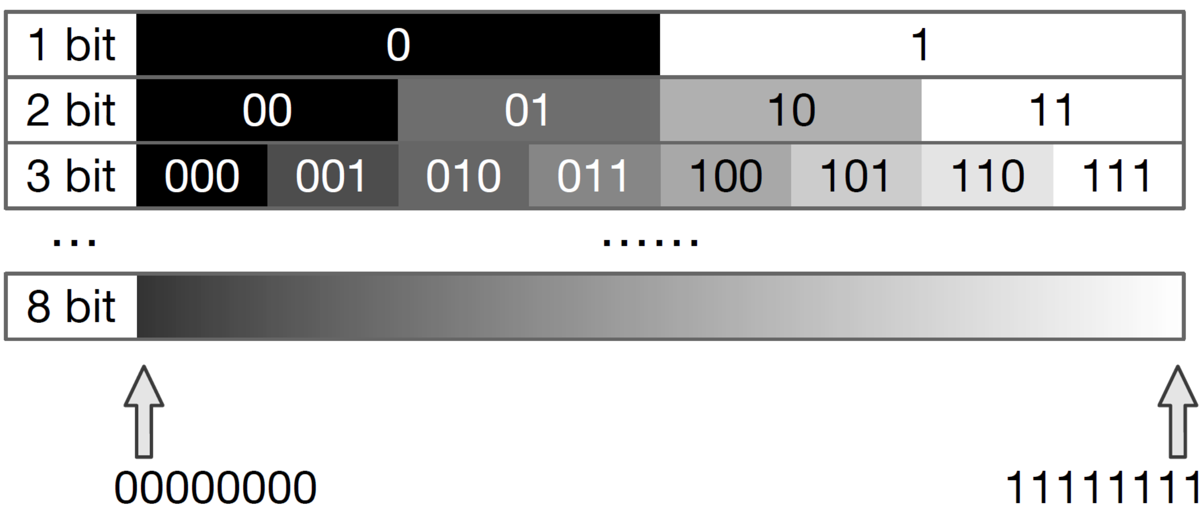
A. Overview of Video Coding

43. Digital video represents video content by sequentially displaying a series of images ("frames") each composed of individual picture elements ("pixels") to produce the illusion of motion.[2] Various metrics are used to describe the quality of video, one of which is its resolution, which refers to the number of pixels making up each frame. Modern digital video commonly has a resolution of 1280x720 ("720p"), 1920x1080 ("1080p") or "4K" which is a term used for the resolutions 4096x2160 and 3840x2160.

44. Each pixel in a frame has a color, which can be represented in terms of its red, green, and blue values ("RGB") or, equivalently, by its luminance, blue chrominance, and red chrominance ("YCbCr"). Each of these components can be expressed as a collection of bits. For example, 8-bit color uses 8 bits for each of the

[1] I refer here to Ex-2012, which contains excerpts from my 2010 book, "The H.264 Advanced Video Compression Standard."
[2] For example, modern digital video commonly uses 30 frames per second.

Y, Cb, and Cr components. A variable represented using n bits can have 2^n possible values. 8-bit color allows for 256 possible values for each component, while 10-bit color allows for 1024 possible values for each component. I created the following diagram to illustrate how different bit depths correspond to more possible values. In these examples, a brightness (luma) value represented with one bit can have two possible values, represented as 0 or 1. A value represented with three bits can have 8 possible values, in this example from 000 (black) to 111 (white). A value represented with 8 bits can have 256 possible values and so on.



45. Storing and transmitting three component values for each pixel in each frame of a digital video would require significant storage capacity and would require extremely fast internet connections in order to be streamed to consumers. As a result, various techniques have been developed for compressing digital video by encoding it to minimize the size of the video for storage and transmission and then decoding it when the video is viewed.

46. Certain common architectures for video coding systems have been established over the years. For example, my 2010 book, "The H.264 Advanced Video Compression Standard," depicts a block diagram of a video encoder and decoder.

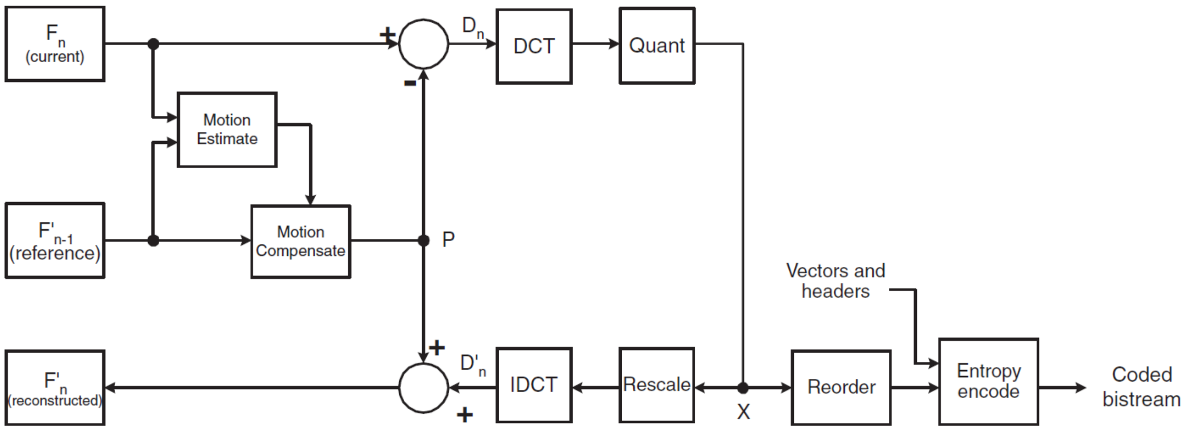


Figure 3.51 DPCM/DCT video encoder

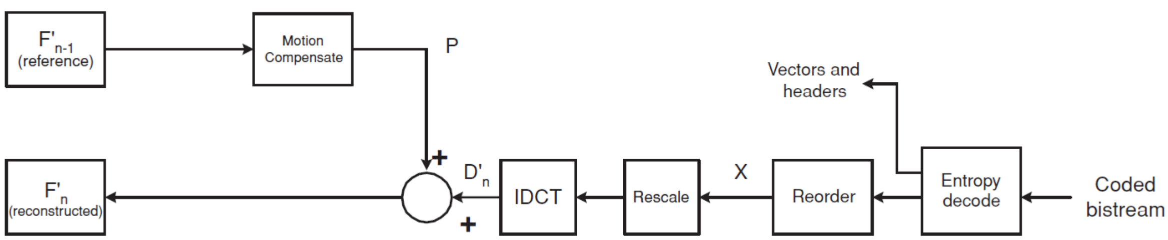


Figure 3.52 DPCM/DCT video decoder

Ex-2012 at page 69.

47. Encoders and decoders work largely symmetrically, with video encoders taking input video frames, $F_{n,}$ and processing these frames in units of a block of pixels. For each block, the encoder chooses a prediction, P, which preferably closely matches the block being processed. The prediction, P, is

subtracted from the block being processed to obtain a residual or difference macroblock D. The residual data is then transformed using a transform such as a Discrete Cosine Transform (the "DCT" block in the figures above) and quantized to eliminate DCT coefficients which contribute negligibly to the residual. Once quantized, the coefficients are reordered and entropy coded to minimize the amount of data written to the output bitstream. Decoders follow a similar process, but in reverse. A coded bitstream is decoded and used to produce a residual block by application of an inverse transform such as an inverse DCT. A prediction, P, is determined using the same method followed by the encoder and combined with the residual to produce a reconstructed block of the current frame.

48. Although various modes of prediction exist in video coding systems, motion compensated prediction is of particular relevance to the '267 Patent. Motion compensated prediction exploits temporal redundancy in video. For example, in a scene in which the camera pans across a cityscape, each frame will be largely similar to the last but shifted in the direction that the camera is moving. Motion compensated prediction allows portions of a current frame to be predicted from reference frames (e.g., a previous frame) so that the encoded residual only accounts for the difference between the current frame and reference frame.

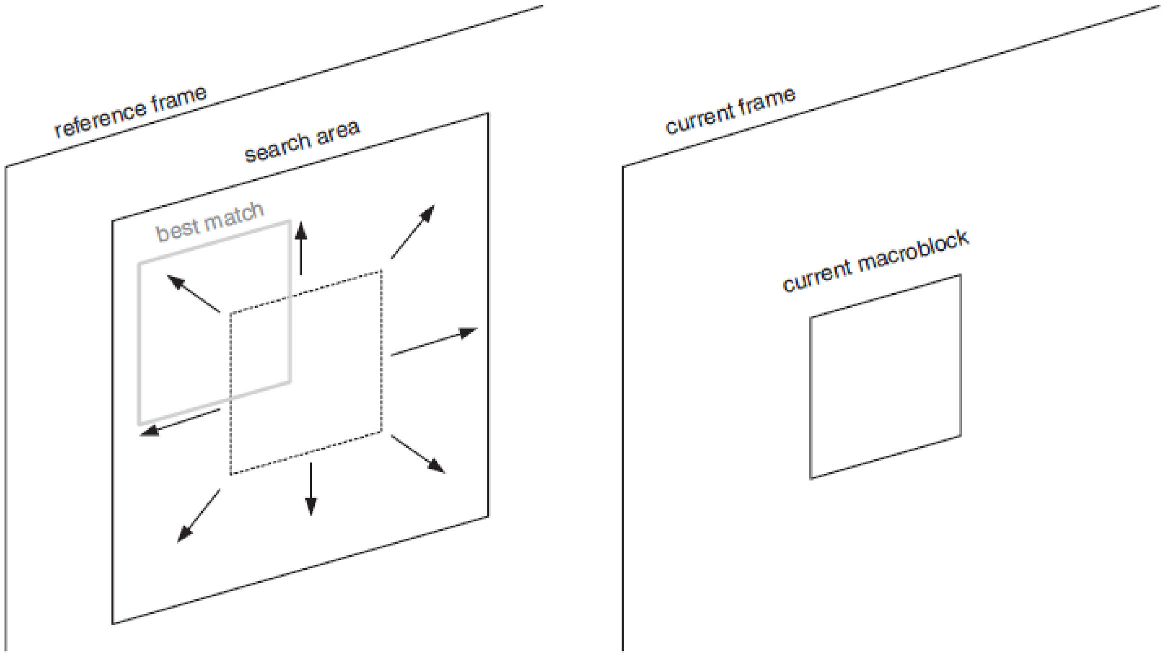
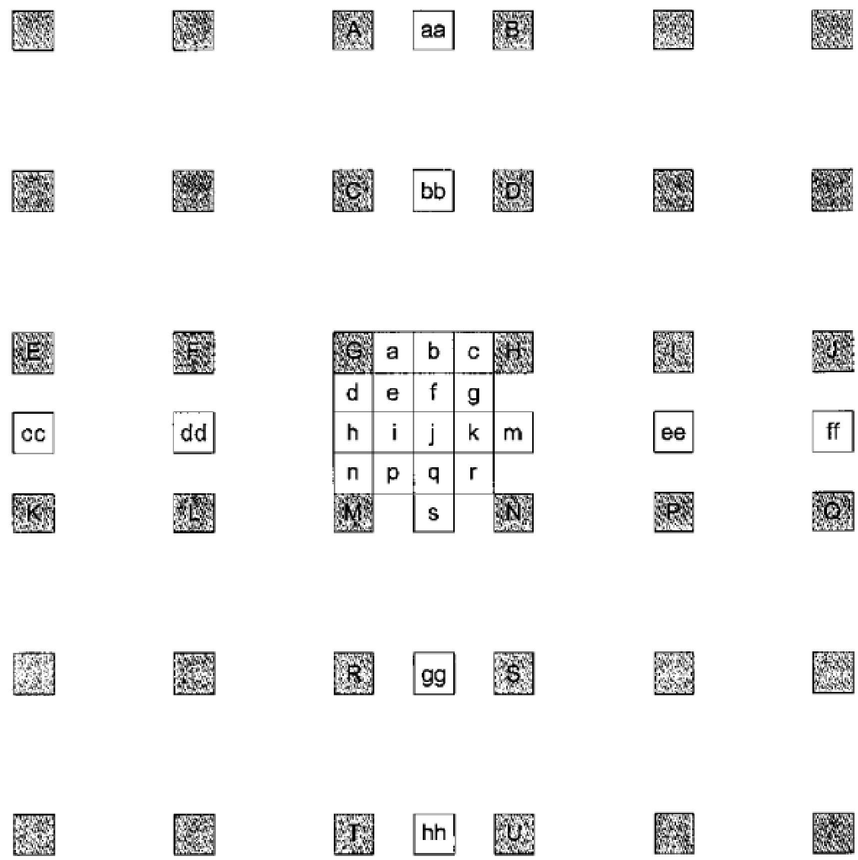


Figure 3.9 Motion estimation

Ex-2012 at page 32.

49. During encoding, motion estimation is used to determine, for a block in the current frame, the best matching block or region of a reference frame. A motion vector represents the displacement between the current block and the matching block or reference block determined during estimation. This motion vector may be encoded in the bitstream or may itself be determined by a motion vector prediction process. During decoding, the process is reversed as described above.



50. In some video codecs, a motion vector may point to a fractional-pixel position in the reference picture, where the fractional-pixel locations refer to locations in between original image pixels. This can enable the encoder and decoder to construct a prediction that is a better match than a prediction that is entirely formed using the actual pixels of the reference picture. The motion-compensated prediction process may involve applying interpolation filter(s)[3] to the image pixels in a

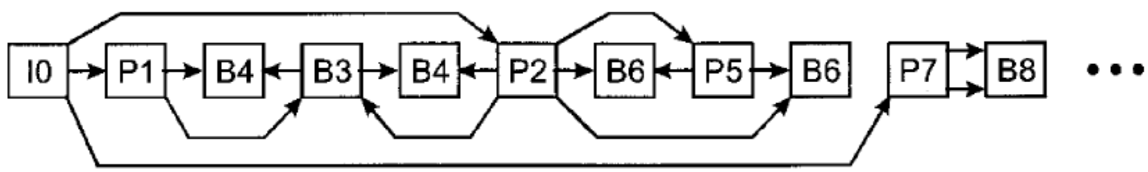
[3] For example, a Finite Impulse Response (FIR) filter such as such as a 6-tap Wiener filter with coefficients corresponding to the weighted contribution of each input integer pixel might be used. Other widths of P-tap filter are also common.

reference picture or frame, to generate interpolated fractional-pixel values. *See* Ex-1001 at 2:60-3:11. Figure 12 of the ‘267 Patent, reproduced above, illustrates examples of image pixel positions (also known as “integer” pixel positions), such as G, H, M and N, and examples of fractional pixel positions such as a, b and c. In the example of Figure 12, fractional pixel positions at ½ and ¼-pixel resolution may be generated. For example, position b is a ½-pixel position between G and H. Positions a and c are ¼-pixel and ¾-pixel positions respectively between G and H. *See* Ex-1001 at 2:60-3:11, Fig. 12.

51. When a sub-pixel motion vector is chosen, generating the prediction block requires interpolation of pixel values in the reference picture to create interpolated values. Interpolation may require a computationally complex series of operations such as multiplications, additions and divisions to arrive at a final interpolated value that can be used as a prediction. Ex-1001 at 3:7-11. Generating potential prediction values for sub-pixel positions is therefore typically computationally expensive.

52. For each current block, an encoder typically selects one matching region in a reference picture, with a corresponding motion vector. However, certain integer-pixel or subpixel values determined using the reference picture may require to be searched for multiple current blocks. For example, potential matching regions that are offset to the right of a first current block may overlap with potential matching

regions that are offset to the left of a second current block, and so on. Thus, an encoder may choose to store interpolated pixel values (i.e., calculated sub-pixel values) for use in motion estimation for multiple current blocks. One strategy that may be computationally efficient and which is used in practical implementations, is for the encoder to calculate all the possible interpolated pixel positions for a reference picture, prior to carrying out motion estimation for blocks of a current picture.



53. The ’267 Patent is directed to an improvement in bi-directional motion compensated prediction or bi-prediction, which is a special case of motion compensated prediction that makes use of two reference blocks and two motion vectors to predict a current block. In bi-prediction, predictions are obtained from each reference block and then combined—for example, by simple averaging or weighted averaging of the two predictions. Figure 8 of the ’267 Patent, reproduced above, illustrates this process. *See* Ex-1001 at 2:35-51, Fig. 8. Each block in Figure 8 represents a picture/frame, and those labeled with a “B” are bi-predicted pictures. The arrows indicate the interdependencies between frames. For example, blocks in B4 may be predicted from P1 and B3. A bi-predicted frame can use references that

are in the past, in the future, or both, relative to the current frame. Bi-prediction often allows a lower residual error to be obtained relative to motion compensated prediction from a single reference, which in turn can lead to improved compression performance. *See* Ex-1001 at 3:25-30.

54. Practical video coding involves a trade-off between decoded image quality, compressed bitrate and computational complexity, as I illustrated in Figure 9.28 of my 2010 book, which I reproduce here:

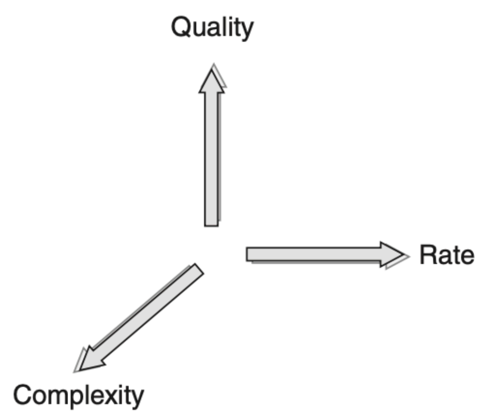


Figure 9.28 Rate, quality and complexity

Ex-2018 at 283.

55. As the '267 Patent explains, a typical video encoder seeks to optimize the trade-off between image quality (or conversely distortion, D) and compressed bitrate (or rate, R) during compression. Ex-1001 at 3:31-48. In other words, a goal of video coding is to provide high compression (i.e., a low bitrate) whilst maintaining good or acceptable image quality (i.e., minimizing distortion). At the same time, it

is important to minimize the computational complexity at both encoder and decoder. For example, many video coding applications such as videoconferencing or video calling, require both encoding and decoding to be carried out in real time on practical devices such as PCs, tablets or cellphones, with limited computational resources. Thus, when designing a practical video codec, an important consideration is to achieve good compression performance (i.e., achieving a low bitrate and a good or acceptable decoded image quality) whilst minimizing computational complexity.

B. Video Coding Standards and Specifications

56. As I explained in my opening report, by 2011, a POSITA would have been familiar with video coding standards such as the ITU-T H.263 and H.264 standards discussed in the ‘267 patent. Ex-1001, 1:34-37. From the late 1980s, the ITU-T and ISO/IEC international standards organizations have developed and published a series of industry standards relating to video coding. These include the well-known ISO/IEC MPEG and ITU-T H.26x standards. Each standard is/was developed in a similar way. A working group of industry experts is instructed by the standards organization (ITU-T and/or ISO/IEC) to seek to develop a new standard. Over a period of typically a few years, members of the working group contribute technology proposals and develop a draft standard which, if approved, is published by the standards organization.

57. Each of these standards specifies a format for a compressed bitstream and a method of decoding such a bitstream to produce a decoded video sequence. These standards do not specify an encoding process. A video decoder conforming to a video coding standard decodes a bitstream that conforms to the specified bitstream format, in order to produce the same output as the specified decoding process. A video encoder encodes a bitstream that conforms to the specified bitstream format and that can be successfully decoded by the specified decoding process.

58. Certain video coding standards have specified methods of bi-directional motion compensated prediction or bi-prediction. These include:

- ISO/IEC MPEG-1 Video and MPEG-2 Video, which each specify methods of bi-prediction of blocks in coded B-pictures, in which a bi-predicted block may be predicted from one past coded I- or P-picture and one future coded I- or P-picture.
- ISO/IEC MPEG-4 Part 2 and ITU-T H.263, which each specify methods of bi-prediction of blocks in coded B-pictures.
- ITU-T H.264 (also published as ISO/IEC MPEG-4 Part 10 and known as Advanced Video Coding or AVC), which specifies more flexible bi-prediction of blocks in coded B-pictures, for example including the ability to predict each block from two coded reference pictures, where

the two coded reference pictures may both be in the past, both in the future or one in the past and one in the future.

59. Certain video coding methods have been developed by technology companies or consortia, including for example Real Video (RV), developed by Real Networks, and Windows Media Video (WMV), initially developed by Microsoft and later published as the VC-1 industry standard.

60. All of the above-mentioned industry standards and proprietary video coding methods share certain common features, including: 1) prediction of image blocks, including the use of motion compensated prediction; 2) transform of residual blocks, e.g., using a Discrete Cosine Transform (DCT) or similar block transform; 3) quantization of transform coefficients; 4) entropy coding of quantized transform coefficients, prediction information (e.g., motion vectors) and other parameters; and 5) complementary encoding and decoding processes.

61. These and other features are illustrated in the following excerpt from my 2010 book (Ex-2012):

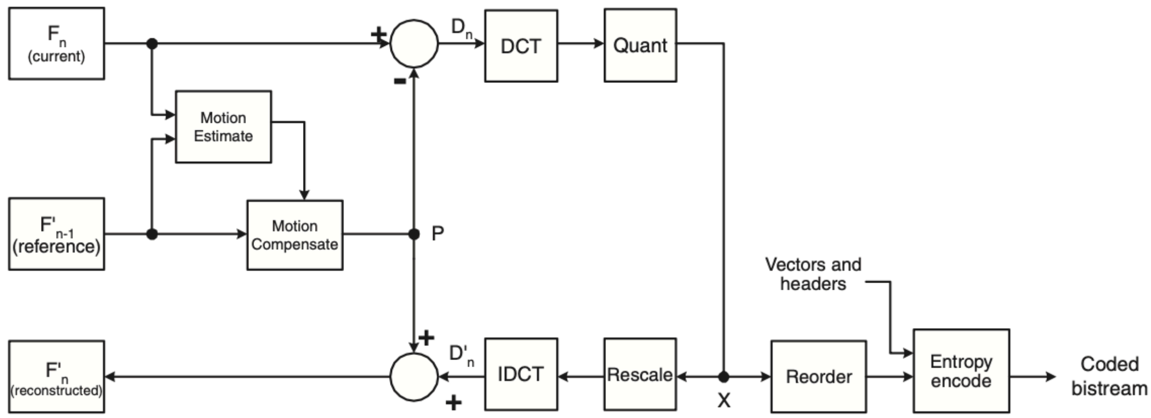


Figure 3.51 DPCM/DCT video encoder

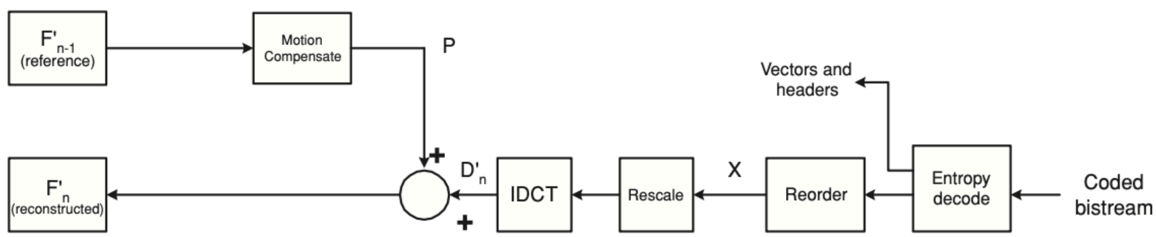


Figure 3.52 DPCM/DCT video decoder

Ex-2012 at 69.

62. As Dr. Freedman recognizes, the block diagrams above illustrate a general architecture for an encoder and decoder using motion compensated prediction which has been around since the 1990s. *See* Ex-1003 at ¶ 3. This general architecture is shared by H.26x-compliant systems, as well as other systems such as Windows Media Video and Real Video (which are referred to in Petitioners' Walker reference). A person of ordinary skill in the art would be accustomed to seeing this general architecture in such video coding systems.

C. The ’267 Patent

63. The ’267 Patent describes improvements to bi-prediction in video coding which increase the accuracy with which bi-predictions are made, and consequently improve the coding efficiency of a video coding system implementing its inventions. *See* Ex-1001 at 3:56-65, 4:29-35. In prior art bi-prediction, after the initial predictions contributing to the final prediction were computed, their values were shifted to the precision of the input reference blocks. *See id.* at 13:43-55, Fig. 10. Figure 10 illustrates the prior art process. To form prediction P1, an integer-pixel or sub-pixel prediction value may be calculated at a higher precision (a higher-precision prediction) using a P-tap filter (Ex-1001 at 12:56-13:8) then a rounding offset is added and the result is bit shifted to the original (input) bit depth to form prediction sample P1. *See* Ex-1001 at 13:8-18. A similar process is applied to form prediction sample P2.

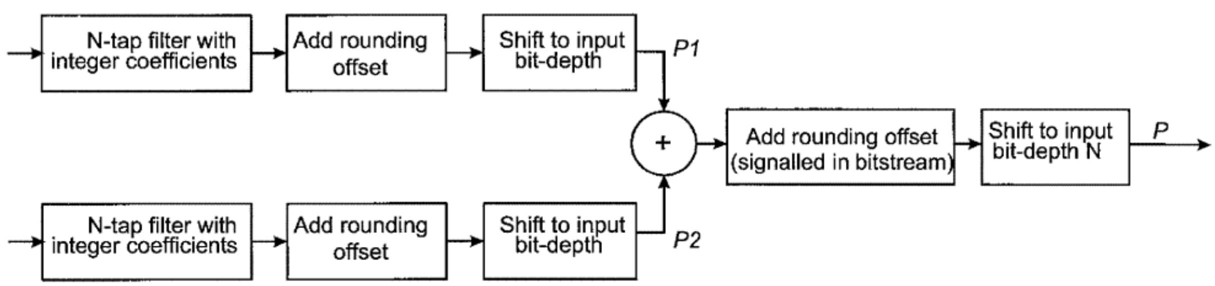


Fig. 10

64. The inventors of the ’267 Patent recognized that in such a prior art implementation, the bi-prediction process was susceptible to the accumulation of

rounding errors which could cause degradation in coding efficiency. *Id.* at 3:56-65. Although other competing solutions to this degradation in bi-prediction efficiency existed, such as adjusting the rounding process itself by alternating upward and downward rounding to reduce rounding error on average, or by adjusting the rounding offset(s), as disclosed in Karczewicz-I (Ex-1005 at ¶¶ 66-69), the inventors of the '267 Patent adopted a technique illustrated in Figure 11. *See id.* at 7:22-24, 13:56-62, Fig. 11. This approach improved the prior art by retaining additional precision resulting from the prediction process until the combination step, lessening the impact of rounding error. *Id.* at 13:43-55.

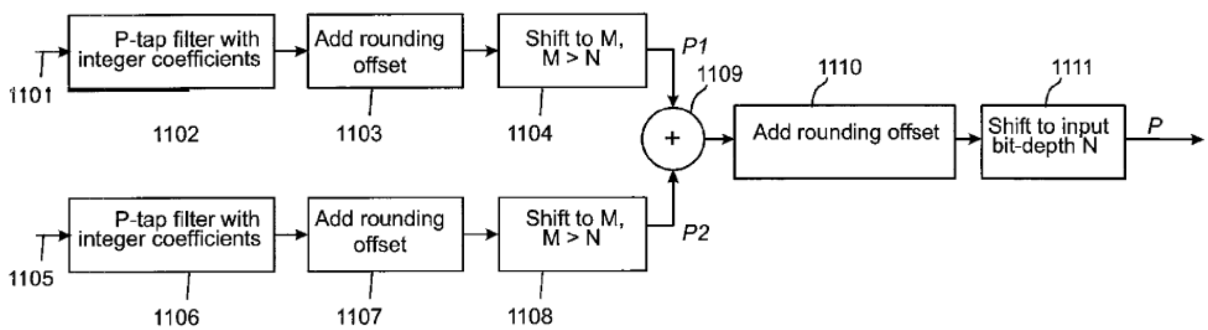


Fig. 11

65. As the '267 Patent explains, when combining two prediction values such as P1 and P2 to form a bidirectional prediction P, in order to predict a current pixel with a certain precision and bit depth, it can be advantageous to calculate P1 and P2 at a higher bit depth and precision than the current pixel. *Id.* at 13:43-55. P1 and P2 are summed at this higher precision and the results are then right-shifted to

create a prediction P that matches the precision of the current pixel. *Id.* at 13:46-48. The advantage of this approach is that rounding errors are reduced. *Id.* at 4:29-31.

1. Predictions in the '267 Patent

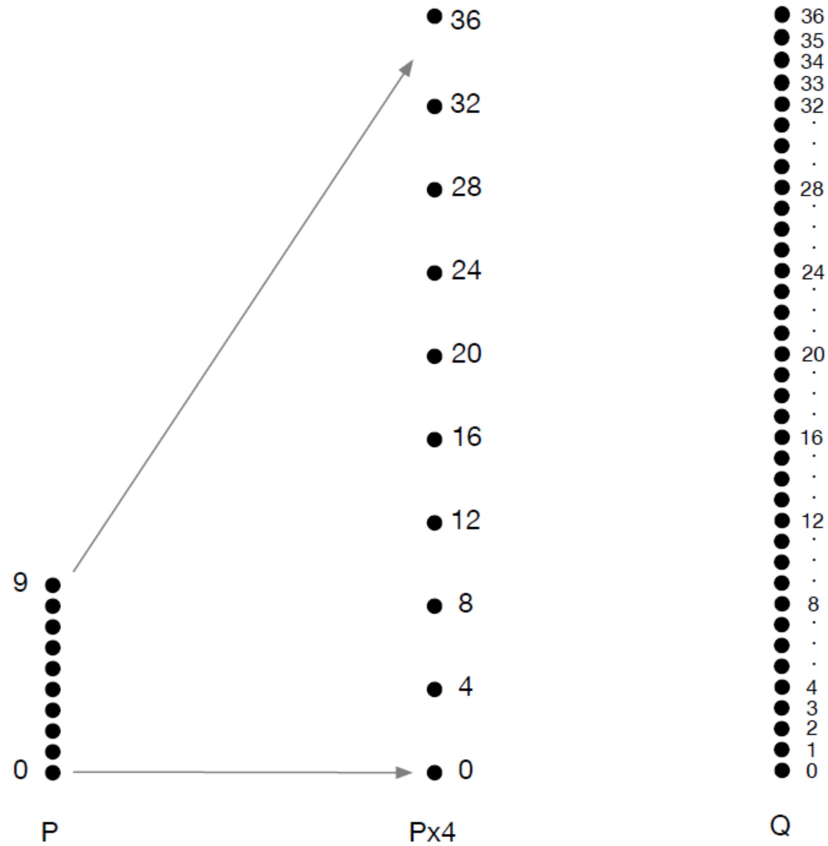
66. In the context of the '267 Patent, a "prediction" for a block of pixels is a value or set of values generated through some coding process that is itself intended to approximately represent the pixel values for that block and that can then be used to form a residual that represents the error between the prediction and its original value. *See* Ex-1001 at 2:1-51 ("in [motion compensated prediction], a prediction for a current frame is formed using a previously coded frame(s), where only the difference between original and prediction signals . . . is encoded."). Because the residual value must be quantized and entropy coded, the efficiency of a prediction method is correlated with the extent to which it minimizes the residual value. *See id.* at 3:56-65.

67. For example, an 8-bit pixel value can have any value in the range 0-255. Given a current pixel value of 120 and a prediction of 116, a residual can be obtained by subtracting the prediction from the current pixel value: 120 – 116 = 4. A 10-bit pixel value can range from 0-1023. Given a current pixel value of 480 and a prediction of 483, the residual is 480 – 483 = -3. In both instances, the aim of the motion compensated prediction process is to minimize the residual value in order to maximize compression.

2. Precision in the ’267 Patent

68. I have reviewed the Declaration and deposition of Petitioners’ expert, Dr. Immanuel Freedman. In my view, the essential problem with his argument with respect to the ’267 Patent’s “second precision, which is higher than said first precision” is that it is based on a logical fallacy. In his Declaration, Dr. Freedman states that a POSITA would understand “precision,” in the context of the ’267 Patent, to refer to “a number of bits ***needed*** to represent ***possible values***.” Ex-1003 at ¶ 39 (emphasis added). I understand Dr. Freedman confirmed during his deposition that this is the only definition of “precision” he is advancing. Ex-2016 at 30:2-9. In the context of the ‘267 Patent, I agree that a POSITA would understand precision in this way. I have emphasized the words “needed” and “possible” to highlight that in my view, Dr. Freedman has incorrectly applied this definition in his analysis. Stated simply, it is ***true*** that a higher precision value will require more bits to represent. It is ***false*** that the use of a higher number of bits determines or implies that a value is at a higher precision.

69. To explain, I refer to an illustration from my opening Declaration, which shows how, given an integer value P in the range $(0, 9)$, the product $P \times 4$ does not have a higher precision than P. *See* Ex-2001 at ¶¶ 104-105.



70. As can be seen, multiplying P by 4 yields values in a new range: $(0, 36)$. Importantly, however, there is a one-to-one mapping between values in the $(0, 9)$ range of P and the $(0, 36)$ range of $P \times 4$. Each unique value of P has a single corresponding value in $P \times 4$, and conversely, every unique value of $P \times 4$ corresponds to a single value of P. This can be restated in terms of Dr. Freedman's proposed definition for "precision": although the ranges of P and $P \times 4$ differ, the number of ***possible values*** is the same, and thus the number of bits ***needed*** to represent P or $P \times 4$ remains unchanged. This is contrasted with Q, which comprises all integers within the range $(0, 36)$, and which includes a further 27 values that are

not possible results of the multiplication $P \times 4$ and which therefore has a higher precision than P.

71. From review of his deposition testimony, it appears to me that, despite stating otherwise in his Declaration, Dr. Freedman's understanding and application of the term "precision" differs significantly from the "number of bits needed to represent possible values." It appears that his true understanding hinges on a statement he makes later in his declaration that "precision" refers to "a number of bits needed to represent the possible values of binary data, ***including uncompressed representations of binary data***." Ex-1003 at ¶ 40 (emphasis added). Dr. Freedman explained during his deposition:

> Q: What I am trying to figure out here from your answer,
> Dr. Freedman, is how the number of bits needed to
> represent possible values changes for a two's compliment
> number?
> A: Oh, it doesn't, and that is the point.
> Q: So it seems that, then, your interpretation of precision
> would apply to a two's complement number and we
> would just look at the greater–
> A: Absolute magnitude.
> Q: Sorry? What was that?
> A: The greater absolute magnitude.

72. Ex-2016 at 33:7-18. Although Dr. Freedman is discussing two's complement numbers here (a way of representing negative numbers in binary form) he is clear that this statement applies to any number. *Id* ("Q: …how [does] the number of bits needed to represent possible values change[] for a two's complement number? A: Oh, it doesn't, and that's the point."). What Dr. Freedman therefore appears to mean when he refers to "precision" is the "absolute magnitude" of a value, not the number of bits ***needed*** to represent it. *See id.* at 38:5-39:22. For example, although a variable that can take only the values of either 0 or 1,000,000 can be represented as a single bit, which is all that is needed to represent two possible values, Dr. Freedman suggests that this would not be an "uncompressed" binary format and therefore not representative of the variable's "precision." *See id.*

73. This understanding is at odds with my own, with the statements that Dr. Freedman makes in his Declaration, and with my view of the understanding of a POSITA. Dr. Freedman's arbitrary reliance on "including uncompressed values" effectively reads "needed" out of the proper understanding of "precision." Instead of considering how many bits are needed to represent a value, Dr. Freedman appears to have defined "precision" to encompass the number of bits an implementation actually uses, regardless of whether or not those bits are actually needed to represent possible values.

74. Dr. Freedman's apparent understanding also implies that an arbitrary 8-bit number would have the same precision before and after it was right-shifted 7 bits and left-shifted 7 bits, even though before the shifts the arbitrary 8-bit number would have 256 possible values and the result after the shift there would be only 2 possible values. A person of ordinary skill in the art would not agree with this interpretation.

75. Considering my example of P and $P \times 4$, Dr. Freedman would appear to contend that $P \times 4$ is actually more precise than P simply because 36 is bigger than 9 (and 32 is bigger than 8, and so on). Therefore, given the same P in the range $(0,9)$, Dr. Freedman appears to take the position that the precision of *any* product $P \times N$ is determined by the magnitude of N. That is, Dr. Freedman appears to be arguing that by setting N to an arbitrarily large number, one can achieve an arbitrarily large precision.

76. A POSITA would be aware that calculating a value such as a prediction at a lower precision then multiplying the prediction by a constant does not and cannot increase precision or in any way improve the quality or accuracy of the prediction. In contrast, calculating a prediction value at a higher precision and therefore at a higher accuracy, can provide a more accurate prediction. *See* Ex-1001 at 4:36-43 (correctly equating higher precision prediction signals with higher accuracy of prediction). The embodiments of the '267 Patent disclose the calculation of prediction values at higher precisions, which in turn results in a higher accuracy

combined prediction. This is because calculating a prediction value at a higher precision leads to a prediction that can take on a larger number of possible values than a prediction value that was calculated at a lower precision. Calculating a prediction value at a lower precision and then scaling it by multiplication with a constant, does not result in a prediction that can take on a larger number of possible values and does not result in higher accuracy of prediction.

77. Dr. Freedman's apparent "absolute magnitude" interpretation of "precision" would in my view not be shared by anyone (such as a POSITA) who has taken an introductory laboratory course at university. In such courses, students are taught that they must account for the actual precision of their measurements by ensuring that their answers do not over-represent how precise they are—a technique known as "significant figures." Dr. Freedman's apparent interpretation of precision is at odds with this. According to Dr. Freedman's interpretation, by simply multiplying each measurement or value by a large enough constant number, an unlimited level of precision is possible. This is clearly not the case, and would lead to absurd results.

78. As an example of the counterintuitive results of Dr. Freedman's apparent understanding of "precision," quantities that clearly differ in their precision would have the same precision due to similarities in absolute magnitude. For example, under such an understanding, quantities rounded at different levels (e.g.,

to the nearest 1, 10, 100, or 1000) would all have the same precision. But this is not the case. For example, a value 8582.2 may be expressed as the numbers 8582, 8580, 8600, or 9000, which each correspond to different levels of precision. Indeed, the process of rounding just illustrated consists of reducing the precision of 8582 to the nearest 1, 10, 100, or 1000. In this example, under Dr. Freedman's approach, each rounded number—8582, 8580, 8600, and 9000—would apparently each have a similar or the same precision due to their closeness in absolute magnitude. A person of ordinary skill in the art would not agree. Using the number 9000 to represent a value 8582.2 clearly involves a different degree of precision than using the number 8582. In fact, the difference in precision relates to the different number of digits needed to represent the value. In the context of predictions in video coding, such a difference in precision can have a direct impact on compression performance and therefore on the video quality obtained for a given bitrate. A less precise prediction (i.e., an increased rounding error) can cause a degradation in coding efficiency. Ex-1001 at 3:56-57.

79. Accordingly, although "precision" can be understood as the "number of bits needed to represent possible values," in the context of the '267 patent, in my view it must not be conflated with the magnitude of a given value.

D. Alleged Prior Art

1. Ground 1 – Walker

80. U.S. Application No. 2005/0281334 (Ex-1004) ("Walker") was published on December 22, 2005. The application describes a method and apparatus for "decoding of multiple weighted bi-directional encoding schemes with a single decoder." Ex-1004 at ¶ 8. Walker aims to design a combined encoder for MPEG-4 Part 2, Real Video 9 ("RV9"), Windows Media Video 9 ("WMV9"), and H.264/AVC[4]. *See id.* at ¶¶ 6, 71-72 (describing "A Single Implementation Satisfying MPEG-4, RV9, WMV9 and H.264").

81. Walker individually describes the operation of conventional bi-prediction in each of the MPEG-4 Part 2 (¶¶ 47-49), RV9 (¶¶ 50-55), WMV9 (¶¶ 56-57), and H.264 (¶¶ 58-70) video coding standards and specifications. Walker explains that MPEG-4 Part 2 only supports bi-prediction in its "Simple Scalable Profile," which includes a mechanism for averaging predictions from two frames but does not have any option to apply weights to the predictions being averaged,

[4] The H.264/AVC standard was co-published as ISO/IEC MPEG-4 Part 10 in 2003. When the authors of Walker refer to "MPEG-4," they are referring to the older (1999) MPEG-4 Part 2 standard. I shall refer to "MPEG-4 Part 2" to distinguish it from the H.264/AVC/MPEG-4 Part 10 standard. MPEG-4 Part 2 and MPEG-4 Part 10 specify very different bitstream formats and decoding methods. An MPEG-4 Part 2 encoder is not compatible with an MPEG-4 Part 10/H.264 decoder and vice versa.

meaning that the two predictions necessarily contribute equally to the final prediction. *Id.* at ¶¶ 47-49. RV9, on the other hand, does provide a mechanism for weighting each prediction, either using weights signaled in the bitstream or derived from a co-located macroblock using direct mode. *Id.* at ¶¶ 50-55. Walker describes WMV9 as having an averaging process that is the same as in MPEG-4 Part 2. *Id.* at ¶¶ 56-57.

82. Walker devotes significantly more attention to H.264 in its specification,[5] describing four different scenarios depending on how weighting is applied to predictions when they are combined into the final prediction. An H.264-compliant decoder determines the weighting scenario based on the value of the predFlagL0 and predFlagL1 variables, which indicate whether the first and second reference frames are subject to weighting, respectively. *Id.* at ¶ 59. In the trivial scenario (predFlagL0=0 and predFlagL1=0), H.264's bi-prediction operates similarly to MPEG-4 Part 2 and WMV-9, simply averaging the two predictions without performing any weighting. *Id.* The "Case 1" and "Case 2" weighted prediction scenarios do not correspond to bi-prediction because only one reference frame contributes to the final prediction. *Id.* at ¶ 66. As Walker explains, Case 1

[5] Walker calls H.264 the "most complicated" of the prior art standards it considers and uses it as the starting point for its unified decoder implementation. Ex-1003 at ¶ 74.

"results in weighted forward P Frame prediction" and Case 2 "results in weighted backward P Frame prediction." *Id.* at ¶¶ 62-66.

83. Walker's H.264 "Case 3" corresponds to H.264 bi-prediction and applies weights w0 and w1 to pred0 and pred1, respectively. *Id.* at ¶¶ 67-70, Eqn. 13. The equation which Petitioners and Dr. Freedman rely on in their Ground 1 arguments corresponds to this final H.264 case.

$$Final\_pred = Clip1[(((pred0)w0 + (pred1)w1 + 2^{\log WD}) >> (\log WD + 1)) + ((o_0 + o_1 + 1) >> 1)] \qquad (13)$$

Id. at ¶ 68, Eqn. 13.

84. Walker provides a description of each variable used in Equation 13 in Table 1, which explains that w0 and w1 are weighting factors applied to, but not correlated with, the prediction values pred0 and pred1.

TABLE 1

Variables	Description	Range	No. of Bits
logWD	The base 2 logarithm of the denominator for all the luma or all the chroma weighting factors (luma and chroma weighting factors can be derived separately)	0 to 7	3
w0	The weighting factor applied to the luma or chroma prediction value for the first reference picture (list "0")	−128 to 127	8
w1	The weighting factor applied to the luma or chroma prediction value for the second reference picture (list "1")	−128 to 127	8

TABLE 1-continued

Variables	Description	Range	No. of Bits
o_0	The additive offset applied to the luma or chroma prediction value for the first reference picture (list "0")	−128 to 127	8
o_1	The additive offset applied to the luma or chroma prediction value for the first reference picture (list "0")	−128 to 127	8

Id. at ¶ 69, Table 1.

85. After describing MPEG-4 Part 2, RV9, WMV9, and H.264 bi-prediction, Walker then begins to describe the construction of a "universal formula" that accounts for the prediction methods used in each. *Id.* at ¶ 71-72. Walker begins by constructing Equation 14, which unifies each scenario for H.264. *See id.* at ¶ 74, Eqn. 14. Table 2 describes the bit widths involved in the operations of Equation 14. *Id.* at ¶ 84, Table 2. Walker continues to develop Equation 14 to account for MPEG-4 Part 2, RV9, and WMV9. After additional modifications, Walker arrives at Equation 18, which is reproduced below. *Id.* at ¶¶ 91-92, Eqn. 18. Walker's Table 4 describes the corresponding bit widths involved in Equation 18. *Id.* at ¶¶ 91-93, Table 4. Petitioners rely on Table 2 for the -626 Petition and Table 4 for the -627 Petition. The difference between the two is that Table 2 includes (pred0)w0 and (pred1)w1 with bit widths of 16, while Table 4 includes (pred0)w_A>>6 and (pred1)w_B>>6 with bit widths of 17. However, the specific bit widths involved is

immaterial to the Walker Ground. As the substance of Petitioners' Walker Ground is the same regardless of whether Table 2 or Table 4 is used, my opinions apply to both. I refer to (pred0)w0 and (pred1)w1 in this Declaration, but my opinions also apply to (pred0)w_A>>6 and (pred1)w_B>>6. That Walker describes bit widths of 17 bits after a right shift in Table 4 for these values rather than 16 bits is immaterial to the issues in dispute.

$$pred_{ij} = ((((pred0)w_A) \gg 6 + ((pred1)w_B) \gg 6) + 2^7) \gg 8 + \text{Offset} \qquad (18)$$

2. Ground 2 – Karczewicz-I and Karczewicz-II

a) Karczewicz-I

86. U.S. Application No. 2011/0007799 (Ex-1005) ("Karczewicz-I") was published on January 13, 2011. It describes conventional bi-prediction as implemented in H.264/AVC in default weighted, implicitly weighted and explicitly weighted forms. Ex-1005 at ¶ 44. Default weighted prediction performs simple averaging of two predictions by applying equal weights, as described by Walker in the trivial H.264 weighting scenario. *See id.* at ¶¶ 44, 60. Implicit weighted prediction and explicit weighted prediction both make use of non-equal weights but differ in how they compute the weights. In implicit weighted prediction, weights are determined based on the temporal distance between reference frames and the current

frame. *See id.* at ¶¶ 45, 61. In explicit weighted prediction, the weights that should be applied to each prediction are explicitly coded in the bitstream. *See id.* at ¶¶ 46, 62. Karczewicz-I sets forth equations for each of these three scenarios.

Default	Bidirectional prediction: $pred(i,j)=(pred0(i,j)+pred1(i,j)\boxed{+1})>>1$
Implicit	Bidirectional prediction: $pred(i,j)=(pred0(i,j)*w0+pred1(i,j)*w1\boxed{+32})>>6$
Explicit	Bidirectional prediction: $pred(i,j)=(pred0(i,j)*w0+pred1(i,j)*w1+\boxed{2^r})>>(r+1)+((o1+o2\boxed{+1})>>1)$

Id. at ¶¶ 60-62, (annotated).

87. Karczewicz-I explains that these bidirectional prediction calculations employ rounding adjustments, generally adding a rounding adjustment of 2^{r-1} prior to a right shift by r binary places and that these frequent rounding operations can reduce the performance of prediction. *Id.* at ¶¶ 63-64. For example, rounding offsets (highlighted) of +1, +32, 2^r and +1 are expressed in the three equations above. Karczewicz-I proposes to mitigate the effect of such a loss of precision due to rounding and to improve the performance of prediction by adjusting the use of rounding offsets.

88. Karczewicz-I's proposed adjustments include: combining multiple offsets, using different offsets for uni- and bi-directional prediction, and adaptively selecting upward, downward or no rounding offsets. *Id.* at ¶¶ 64-68.

89. Thus, Karczewicz-I proposes to improve the performance of bidirectional prediction processes by manipulating the offsets used in the rounding process. Importantly, Karczewicz-I does not propose any change in the precision of either of the two prediction values used in bidirectional prediction. Karczewicz-I incorporates the disclosure of Karczewicz-II as I explain later.

b) Karczewicz-II

90. U.S. Application No. 2009/0257499 (Ex-1006) ("Karczewicz-II") was published on October 15, 2009. Unlike Karczewicz-I, Karczewicz-II is not directed to bi-prediction. Instead, this reference concerns subpixel interpolation. *See, e.g.,* Ex-1006 at ¶ 7. Because motion vectors used for motion-compensated prediction can point to fractional-pixel values, the effective value of pixels at those fractional locations (i.e. subpixels) are obtained by interpolation, as I explain in the "Overview of Video Coding" background section above. To interpolate subpixel values, video codecs typically use linear filters such as a 6-tap Wiener filter with coefficients corresponding to the weighted contribution of each input integer pixel. *See id.* at ¶ 68. For example, half-pixel values can be interpolated from integer pixels as shown in Figure 4B. *Id.* at ¶ 70.

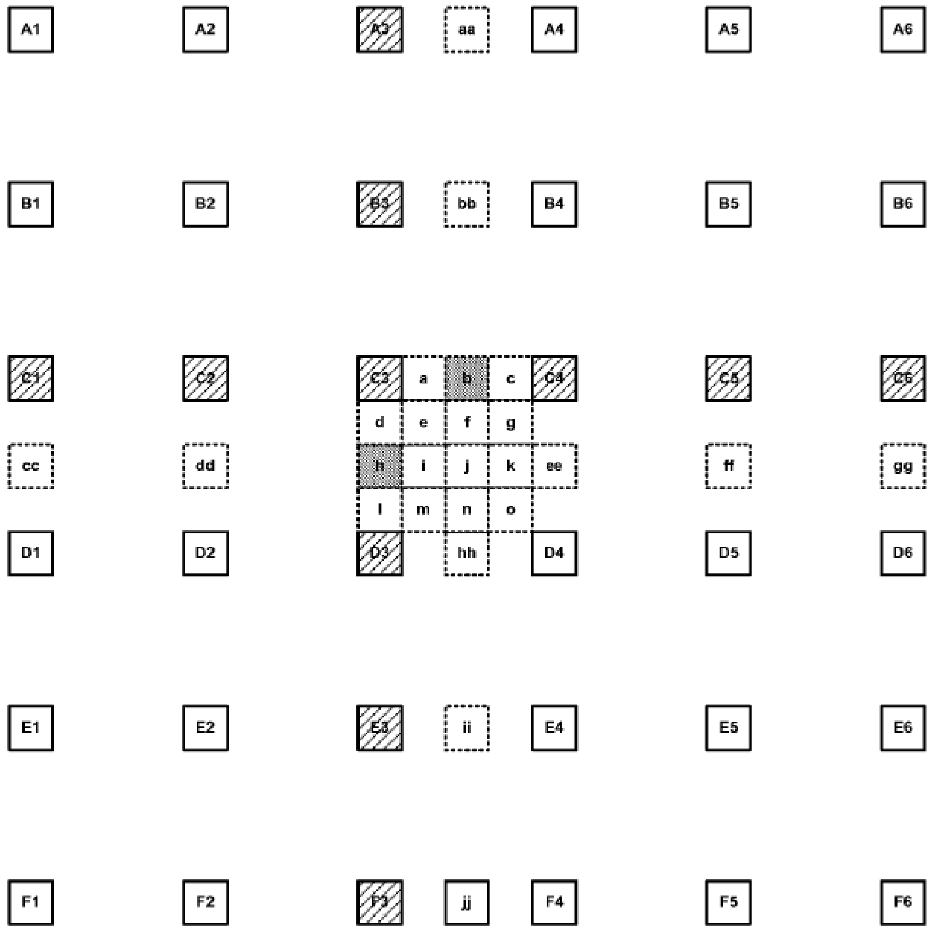


FIG. 4B

91. Using the 6-tap filter weights of [1, -5, 20, 20, -5, 1] described in Karczewicz-II, half-pixels "b" and "h" can be computed from the shaded integer pixels as discussed in paragraph 70:

$$b = C1 - 5 \cdot C2 + 20 \cdot C3 + 20 \cdot C4 - 5 \cdot C5 + C6$$

$$h = A3 - 5 \cdot B3 + 20 \cdot C3 + 20 \cdot D3 - 5 \cdot E3 + F3$$

92. Similarly, quarter-pixel values can be computed from both half-pixel and integer pixel values, as illustrated in Figure 4D. *See id.* at ¶ 72-74. Because of this dependency on half-pixel values as well as integer-pixel values, quarter-pixels cannot be computed until half-pixel values are known.

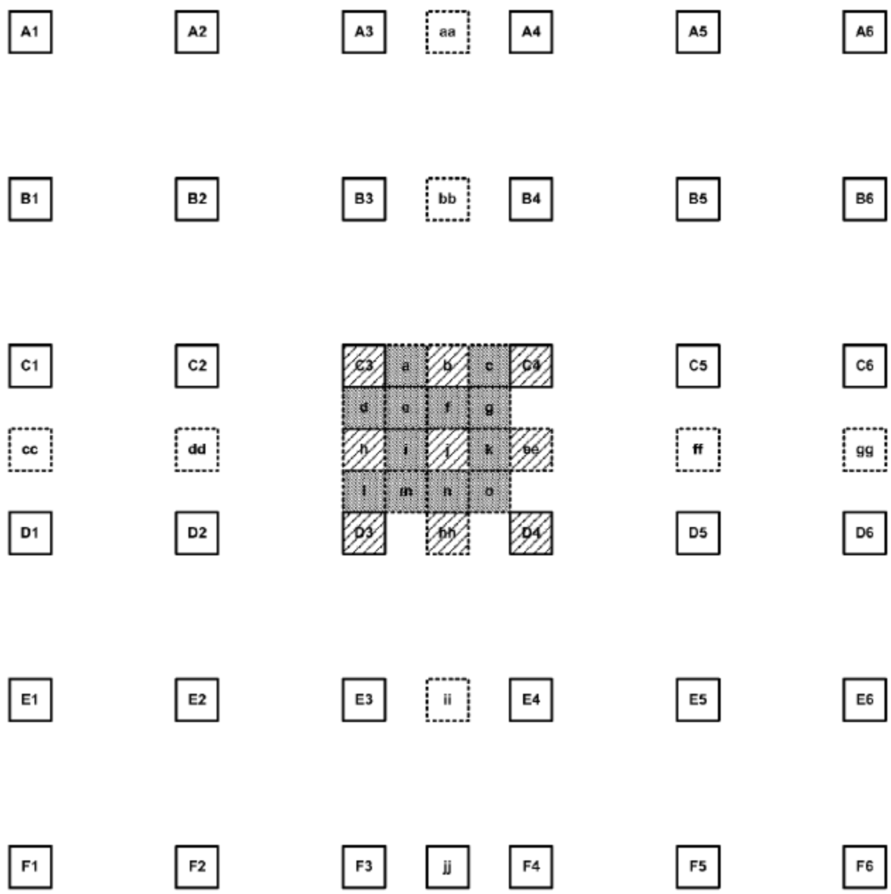


FIG. 4D

93. Karczewicz-II notes that H.264/AVC interpolation is performed with intermediate rounding operations, which round half-pixel values down to the input

precision before they are used to compute quarter-pixel values. *Id.* at ¶ 86. The reference suggests that, instead of rounding half-pixel values prior to quarter-pixel interpolation, the interpolated half-pixel values should be maintained at a high precision (e.g., 16 bits) until quarter-pixel interpolation has concluded, at which point they are rounded to the input precision (e.g., 8 bits). *See id.* at ¶¶ 100-102, 104, Tables 5-7.

VII. Petitioners' grounds do not anticipate or render obvious the challenged claims.

A. Ground 1 – Walker

1. Independent Claims 1, 7, and 13

94. There are two fundamental errors that Petitioners make in their Walker ground: 1) Petitioners misidentify the product of a prediction and a weight as being a "prediction" and 2) Petitioners make use of a flawed understanding of "precision."

a) Neither (pred0)w0 nor (pred1)w1 is a "prediction" as claimed in the '267 Patent.

95. Walker describes a process for calculating a weighted prediction in the context of prior art standards such as H.264/AVC, MPEG-4 and Windows Media Video 9/WMV9. Ex-1004 at ¶ 6. Each of these standards includes methods for bi-prediction, in which predictions are obtained from each of two reference frames and combined to yield a final prediction. *Id.* Walker's Figure 5 (reproduced below) illustrates a generic weighted bi-prediction process. *Id.* at Fig. 5, ¶ 14, 34.

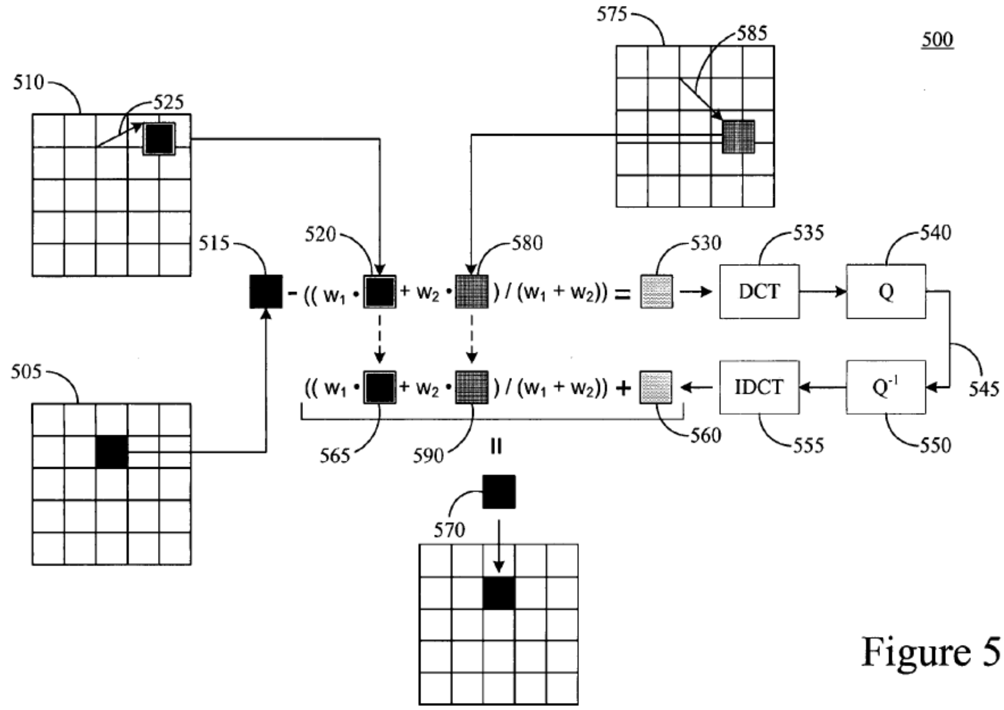


Figure 5

96. During weighted bi-prediction, two predictions (520 and 580 above) are each multiplied by weighting factors w_1 and w_2 respectively. The products of this multiplication are then added, and the sum is divided by the normalization factor $w_1 + w_2$, which ensures that the contribution of the two predictions to the final prediction sums to 1. *Id.* at ¶ 36. Because division is a complex operation in digital signal processing, as Walker explains, it is often preferable to include the normalization factor in the weights w1 and w2 rather than performing division after summing the weighted predictions. *Id.* at ¶¶ 35-36. This is equivalent to distributing the $\frac{1}{w_1+w_2}$ normalization factor over the parenthesis, yielding modified weights $w_1' = \frac{w_1}{w_1+w_2}$ and $w_2' = \frac{w_2}{w_1+w_2}$. *See id.* Regardless of whether the normalization factor is included

in the weights or included by division later, the modified weights w_1' and w_2' represent the actual contribution of each prediction to the final prediction.

97. Although weights may be signaled directly in an encoded bitstream, Walker's Figure 6 illustrates another technique, called "direct mode" in H.264, by which the weights w_1 and w_2 may be derived.[6] *Id.* at Fig. 6, ¶ 38. In direct mode, no weights are signaled at all, and their values are instead derived based on the temporal distance between the two reference frames used for bi-prediction. Ex-1004 at Fig. 6, ¶¶ 15, 38. I have reproduced Figure 6 below.

[6] Walker uses the notation w_1 and w_2, or w_0 and w_1, to refer to the two weighting factors employed during H.264 bi-prediction. Regardless of the notation used, the concepts are described consistently throughout Walker.

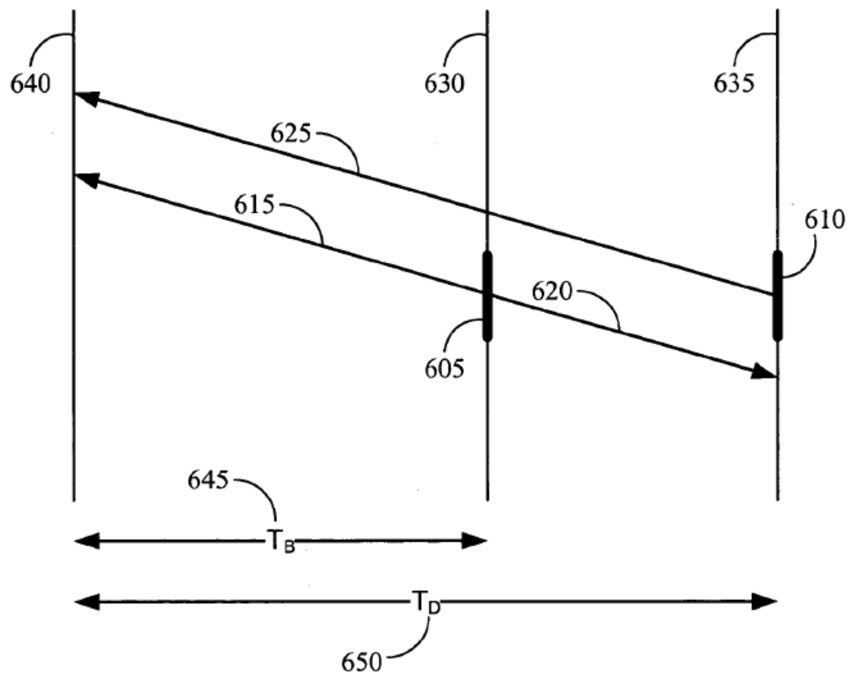


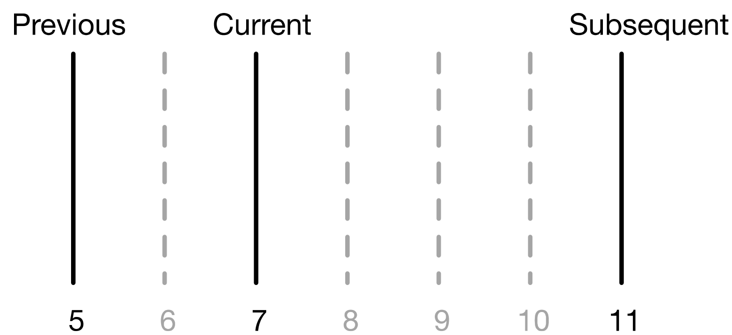
Figure 6

98. There are three relevant frames illustrated in Figure 6: the current frame (630), the subsequent reference frame (635) and the previous reference frame (640). *Id.* As Walker explains, the value T_B is computed as the temporal difference (i.e., the difference in frame count) between the previous reference frame and the current reference frame. *Id.* at ¶ 38-39. Similarly, T_D represents the total distance between reference frames, and thus is computed as the temporal difference between the previous reference frame and the subsequent reference frame. *Id.* Using these values, the weights for each frame can be computed without their need to be signaled, as shown in Walker's equations 3 and 4.

$$w_1=(T_B/T_D) \qquad (3)$$

$$w_2=((T_D-T_B)/T_D) \qquad (4)$$

99. It is apparent, from examination of Figure 6, that T_D must always be larger than T_B. Accordingly, each of the weights w_1 and w_2 is less than 1. Each of the weights is based on the temporal distance between frames and so does not correlate with the prediction values pred0 and pred1. To better illustrate this, I have provided my own version of Figure 6, with exemplary values added for each frame number.



100. In this example, TB = 7 – 5 = 2, and TD = 11 – 5 = 6. Simple arithmetic yields the following values for the weights:

$$w_1 = \frac{T_B}{T_D} = \frac{2}{6} = \frac{1}{3}$$

$$w_2 = \frac{T_D - T_B}{T_D} = \frac{6-2}{6} = \frac{2}{3}$$

101. The same principle holds for the normalized weights discussed earlier in this section. Because $w_1 + w_2$ will always be a larger value than w_1 or w_2

individually, it follows that the normalized weights w_1' and w_2' must also be less than 1. Walker's weights are each either (a) two fractional values that add up to 1 or (b) two scaled fractional values that add up to a power of 2. *See id.* at ¶¶ 36-37. Taking the example above, where $w_1 = \frac{1}{3}$ and $w_2 = \frac{2}{3}$, scaled fractional weights could be used instead by applying a scaling factor of $2^4 = 16$ such that w0 and w1 add to 16, with the factor of 2^4 compensated for by a bit shift:

$$w0 = round\left(\frac{1}{3} \cdot 16\right) = 5$$

$$w1 = round\left(\frac{2}{3} \cdot 16\right) = 11$$

102. Considering example prediction samples pred0 = 121 and pred1 = 116, with a current pixel value of 120, Walker's combined prediction would be:

$$Pred = \big((5 \cdot pred0) + (11 \cdot pred1) + 8\big) \gg 4 = 118$$

This combined prediction *pred* is an approximation of the current pixel value as it is composed of two prediction values, each weighted by fractions that sum to 1 (or a bit-shifted version of 1) and would yield a residual of 120 – 118 = 2. The combined prediction *pred* can be used to predict the pixel value of 120. Crucially, neither (pred0)w0 nor (pred1)w1 are predictions of a current pixel value. Calculating each product individually and right-shifting by 4 to compensate for the scaling factor yields values that are ***not*** representative of the current pixel value.

$$(w0 \cdot pred0) \gg 4 = 605 \gg 4 = 37$$

$$(w1 \cdot pred1) \gg 4 = 1276 \gg 4 = 79$$

Neither of these are predictions for the current pixel value of 120, regardless of any bit shifting. Each would result in a prediction error that is around the same order of magnitude as the pixel value itself. A prediction, i.e., a value that can be used to predict the current pixel value, is obtained only when the two products (pred0)w0 and (pred1)w1 are added together (with optional bit-shifting). Continuing with the same example, (pred0)w0 + (pred1)w1 = 605 + 1276 = 1881. After adding a rounding factor and right-shifting by 4, the final prediction is 117, which is an approximation of the current pixel value of 120.

103. For these reasons, it is my opinion that Walker's (pred0)w0 and (pred1)w1 are not "predictions" as claimed in the '267 Patent. These values are at most intermediate values used in the calculation of an actual prediction (*pred*). Accordingly, I disagree with Dr. Freedman's opinions as to the validity of the challenged claims in view of Walker and hold the opinion that these claims are valid.

b) Neither (pred0)w0 nor (pred1)w1 has a "second precision" which is higher than that of the reference blocks.

104. As stated above, in my opinion Dr. Freedman's analysis of "precision" is logically flawed. This error forms the basis of his assertion that each of (pred0)w0 and (pred1)w1 have a "second precision, which is higher than said first precision". By incorrectly applying his apparent understanding that "precision" is

interchangeable with "absolute magnitude," Dr. Freedman draws the erroneous conclusion that (pred0)w0 and (pred1)w1 each have a higher precision than pred0 and pred1 respectively. Even though Walker chooses to store these product values in 16 bits, only 8 bits are ***needed*** to represent their possible values, since each product has only 256 possible values.

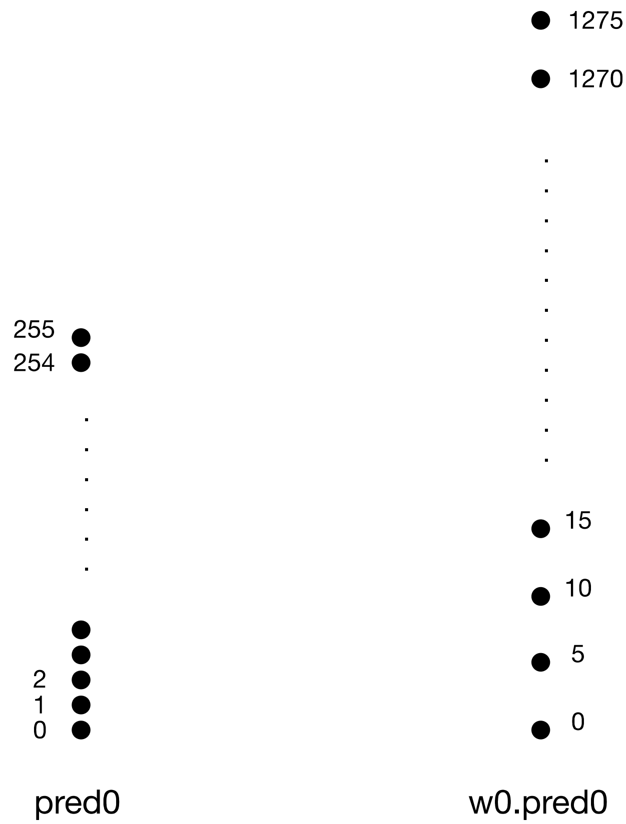
105. Walker describes pred0 and pred1 as each having 8 bits of precision[7], meaning that each can take one of 256 possible values in the range $(0, 255)$. *See id.* at ¶ 60. However, each of the products (pred0)w0 and (pred1)w1 can still only take one of 256 possible values, mapped over a different range. Using the same exemplary value above, w0 = 5, w1 = 11, each product can take one of the following values, where there are exactly 256 possible values in each case, spaced at intervals of 5 or 11 respectively:

$$(pred0)w0 \in \{0, 5, 10, 15, 20, 25, \dots, 1270, 1275\}$$

$$(pred1)w1 \in \{0, 11, 22, 33, 44, 55, \dots, 2794, 2805\}$$

I illustrate this relationship graphically, for the product (*pred*0)*w*0:

[7] H.264 specifies scenarios in which pixel values can have larger bit depths, for example 10 bits per pixel. However, the same reasoning would apply, i.e. a 10-bit pixel value scaled by a constant weighting factor, results in a scaled value that has the same precision as the 10-bit pixel value but a larger range.



106. This reveals two flaws in Dr. Freedman's theory. First, Dr. Freedman incorrectly contends that, because the magnitude of $pred0 < (pred0)w0$, the precision of the product is increased to the absolute magnitude of (pred0)w0. However, as I explain in my discussion of "precision" above, this applies an incorrect understanding of "precision." The product value ($pred0)w0$ can still take only 256 possible values and so can still be represented by 8 bits, with a different 8-bit binary string representing each possible value, i.e., its precision is 8 bits. Although an implementation ***could*** choose to use 16 bits to represent (pred0)w0 or (pred1)w1, as shown in Walker's Table 2, only 8 bits are ***needed***.

107. Second, under Dr. Freedman's incorrect theory of "precision" as "absolute magnitude," (pred0)w0 and (pred1)w1 may not even have the same precision as each other. This is because the maximum value of $255 \cdot w_N$ changes depending on what weight has been chosen. But the '267 Patent describes a "first prediction" and a "second prediction," both of which have the same "second precision, which is higher than said first precision." *See, e.g.,* Ex-1001 at 21:22-27. The only time (pred0)w0 and (pred1)w1 would have the same "greater absolute magnitude," as Dr. Freedman explained in his deposition, would be when w0 = w1. *See* Ex-2016 at 33:7-18. When w0 = w1, the weighted prediction process of H.264 is reduced to simple averaging as shown in Walker's Equation 8:

$$Final\_pred=(pred0+pred1+1)>>1 \qquad (8)$$

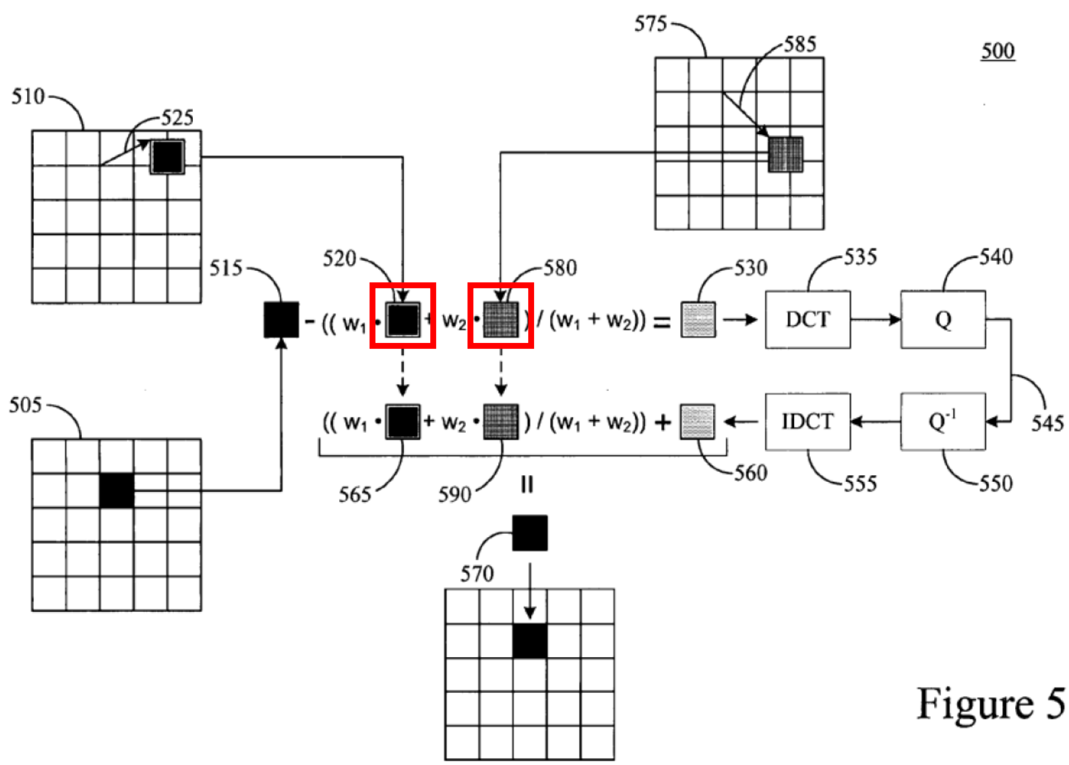
Ex-1004 at ¶ 59, Eqn. 8.

108. However, Dr. Freedman relies on Walker's Equation 13 where $w0 \neq w1$. Accordingly, even if Dr. Freedman were correct that "precision" is equivalent to the greatest "absolute magnitude," Walker would not teach a first and second prediction having a "second precision" as claimed in the '267 Patent.

c) Dr. Freedman's identified "first reference block" and "second reference block" do not contain "pixels" as required by the claims.

109. The challenged claims recite "[determine/determining], for a current block, a first reference block based on a first motion vector and a second reference

block based on a second motion vector, wherein the pixels of the current block, the first reference block, and the second reference block have values with a first precision." Ex-1001 at Claims 1, 7, 13, 19, 25, and 31. Dr. Freedman has incorrectly argued that the "reference block" limitations are satisfied by Walker's "best matching macroblocks." *See* Ex-1003 at ¶ 68 (citing Ex-1004, ¶ 34, Fig. 5). They are not, because Walker's "best matching macroblocks" are predictions which—when determined according to a fractional motion vector—do not consist of pixels. Additionally, several dependent claims, addressed below, require that fractional motion vectors be used.



Ex-1004 at Fig. 5 (annotated).

110. Figure 5 of Walker, reproduced above, illustrates the two "best matching macroblocks" (520 and 580) which Dr. Freedman identifies with the "first reference block" and "second reference block" respectively. Ex-1003 at ¶ 68. I have annotated Figure 5 to show the two identified macroblocks.

111. The claims also recite "wherein the pixels of the current block, the first reference block, the first reference block, and the second reference block have values with a first precision." With respect to these limitations, Dr. Freedman cites to two separate portions of Walker, the first describing a "macroblock" and the second describing the pred0 and pred1 "prediction blocks." Ex-1003 at ¶ 73 (citing Ex-1004 at ¶¶ 30, 60). Dr. Freedman does not address that a "macroblock" as in Walker's paragraph 30 is different from the "prediction blocks" discussed in paragraph 60.

112. Walker explains that "a macroblock is made up of 16x16 pixels" which represent the luminance and chrominance values of a "picture" in the digital video being encoded. *See* Ex-1004 at ¶ 30. However, pred0 and pred1 are not composed of pixel values from a picture (at least when determined according to a fractional motion vector). Instead, "pred0 and pred1 are 8-bit luminance and chrominance . . . samples from prediction blocks from the two reference frames." *Id.* at ¶ 60. Walker explains the process of forming predictions using pixel interpolation with reference to Figure 9, which shows "half pixel interpolation where one interpolated pixel is located between each of the original integer pixels." *Id.* at ¶ 114.

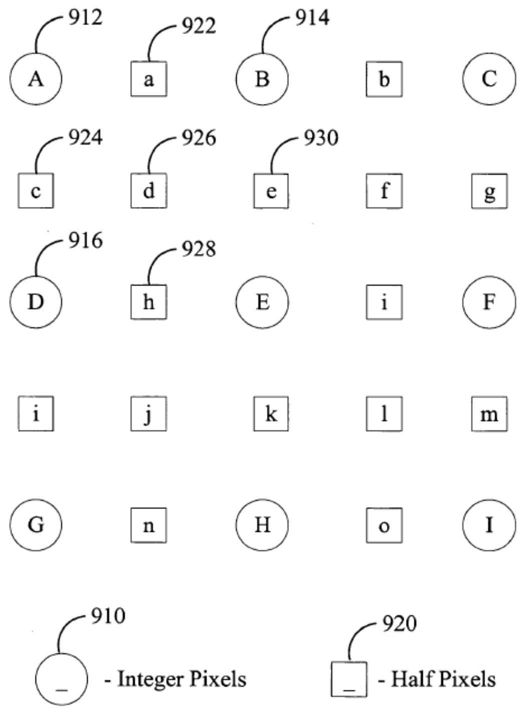


Figure 9

113. As Figure 9 illustrates, the "original integer pixels," which are part of a picture in the video sequence, are filtered to obtain "half pixels." Half pixels represent color values in a prediction block but are not pixels from a picture. Accordingly, Walker cannot disclose that "the pixels of the current block, the first reference block, and the second reference block have values with a first precision" as required by the challenged claims.

2. Dependent Claims 2, 8, and 14

114. Dependent claims 2, 8, and 14 depend on independent claims 1, 7, and 13 respectively. These claims include the additional limitation "wherein in an instance in which said first motion vector points to a subpixel, said first prediction

is obtained by interpolation using pixel values of said first reference block." It is my opinion that these claims are valid for the same reasons discussed above with respect to the claims on which they depend. In addition, I disagree with Dr. Freedman's arguments with respect to these claims because he fails to show that "said first prediction is obtained by interpolation using pixel values of said first reference block."

115. As I discuss above, Dr. Freedman identifies Walker's "(pred0)w0" with the '267 Patent's "first prediction" and "pred0" with the '267 Patent's "first reference block," respectively. Ex-1003 at ¶¶ 76-77. However, with respect to the present claims, Dr. Freedman discusses how "Walker teaches obtaining a prediction for a subpixel" from a reference block. Ex-1003 at ¶¶ 103-104. Regardless of Dr. Freedman's opinions on what constitutes a "prediction," the portions of Walker that he cites only explain that pred0 can be obtained using pixel values of a reference block using interpolation. *See* Ex-1004 at ¶¶ 111, 114. What Dr. Freedman does not explain is how his alleged "prediction," (pred0)w0, can be obtained from pred0 by interpolation—this is because it cannot. (pred0)w0 is the product of pred0 multiplied by a weight, w0. *See id.* at ¶ 68, Table 1. Although Walker discusses the use of a 2-tap FIR filter to obtain subpixel values, no such process is applied to pred0 in order to obtain (pred0)w0. *Compare id.* at ¶ 114 *with id.* at ¶ 68, Table 1. Instead, Walker

teaches using simple multiplication by a constant, w0, to scale pred0 to the weighted value (pred0)w0. *Id.*

116. Accordingly, because the claims expressly require "interpolation," it is my opinion that Dr. Freedman is incorrect in his opinion that Walker discloses claims 2, 8, and 14.

3. Dependent Claims 3, 9, and 15

117. Dependent claims 3, 9, and 15 depend on claims 2, 8, and 14 respectively. These claims include the additional limitation "wherein said first prediction is obtained by interpolation using values of said first reference block by: right shifting a sum of a P-tap filter using values of said first reference block." It is my opinion that these claims are valid for the same reasons discussed above with respect to the claims on which they depend. In addition, I disagree with Dr. Freedman's arguments with respect to these claims because he fails to show the use of a "P-tap filter" to obtain a "first prediction" using "values of said first reference block."

118. As I explain above, Dr. Freedman's alleged "first prediction," (pred0)w0, is obtained from his alleged "first reference block," pred0, by multiplication. *See* Ex-1003 at ¶¶ 76, 85. Multiplication is not a form of interpolation, much less a P-tap filter. Similarly, no right shift is involved in the

multiplication by w0. Accordingly, it is my opinion that Dr. Freedman is incorrect and that claims 3, 9, and 15 are valid.

4. Dependent Claims 4, 10, and 16

119. Dependent claims 4, 10, and 16 depend on claims 2, 8, and 14 respectively. These claims include the additional limitation "wherein in an instance in which said second motion vector points to an integer sample, said second prediction is obtained by shifting values of said second reference block to the left." It is my opinion that these claims are valid for the same reasons discussed above with respect to the claims on which they depend.

120. Dr. Freedman points out that (i) there are instances in Walker in which a motion vector points to an integer sample and that (ii) there are instances in Walker in which a weight such as w1 happens to have a value that is a power of 2. Ex-1003 at ¶¶ 114-115. However, in Walker, these instances are coincidental and there is no causal relation between them. Walker's disclosure does not require that w1 be a power of two value when a motion vector points to an integer sample. Further, Walker does not disclose that a multiplication by w1, where w1 is a power of two, is obtained by shifting values to the left. Thus, Dr. Freedman effectively analyses the additional limitation as if it were: "in an instance in which said second motion vector points to an integer sample, said second prediction ***could conceivably be*** obtained by shifting values of said second reference block to the left."

121. Further, I do not agree with Dr. Freedman that it would be obvious to implement a multiplication by w1 as a left-shift in this situation. *See* Ex-1003 at ¶ 115. Walker's decoder does not know in advance what value of w1 will be applied for a current block. Since w1 can take a number of possible values, including values that are not a power of two, in my view it would not be obvious or computationally efficient to implement a special, extra test for a power of two value of w1, as I illustrate in the following pseudocode:

```
obtain w1
if (w1 is a power of two)
    calculate pred1(w1) using a left-shift
else (w1 is not a power of two)
    calculate pred1(w1) using a conventional multiplication
end
```

122. In my opinion, the obvious implementation is to simply calculate (pred1)w1 using a conventional multiplication, since the extra test (of whether or not w1 is a power of two) may well negate any advantages of carrying out some multiplications as left-shifts, as suggested by Dr. Freedman. Dr. Freedman's rationale that "left-shifting . . . can be performed without the need for a multiplication unit" (Ex-1003 at ¶¶ 115), does not make sense, because a multiplication unit is still

needed for values of w1 that are ***not*** powers of two, and thus ***is*** needed for Walker's calculation of (pred1)w1.

5. Dependent Claims 5, 11, and 17

123. Dependent claims 5, 11, and 17 depend on claims 1, 7, and 13 respectively. These claims include the additional limitation wherein decreasing the precision of the combined prediction by shifting bits of the combined prediction to the right further comprises inserting a rounding offset to the combined prediction before said decreasing. It is my opinion that these claims are valid for the same reasons discussed above with respect to the claims on which they depend.

6. Dependent Claims 6, 12, and 18

124. Dependent claims 6, 12, and 18 depend on claims 1, 7, and 13 respectively. These claims include the additional limitation "wherein the first precision indicates a number of bits needed to represent the values of the pixels, and the second precision indicates the number of bits needed to represent values of said first prediction and values of said second prediction." It is my opinion that these claims are valid for the same reasons discussed above with respect to the claims on which they depend. Further, should the Board reject the parties' definition of precision, I note that this dependent claim expressly limits the meaning of "precision" to "the number of bits needed to represent values of said first prediction and values of said second prediction," without additional qualification. Accordingly,

my arguments with respect to the independent claims are applicable to dependent claims 6, 12, and 18, which expressly specify the meaning of “precision” to be the same as my own.

7. Independent Claims 19, 25, and 31

125. Independent claims 19, 25, and 31 are decoding claims corresponding to encoding claims 1, 7, and 13. I understand that Dr. Freedman’s opinions with respect to these claims are largely the same as the ones he offers for their encoding counterparts, and so I disagree with him for the same reasons.

8. Dependent Claims 20, 26, and 32

126. I note that Dr. Freedman’s opinions with respect to decoding claims 20, 26, and 32 are substantially the same as his arguments with respect to encoding claims 2, 8, and 14. Accordingly, it is my opinion that these claims are valid for the same reasons.

9. Dependent Claims 21, 27, and 33

127. I note that Dr. Freedman’s opinions with respect to decoding claims 21, 27, and 33 are substantially the same as his arguments with respect to encoding claims 3, 9, and 15. Accordingly, it is my opinion that these claims are valid for the same reasons.

10. Dependent Claims 22, 28, and 34

128. I note that Dr. Freedman's opinions with respect to decoding claims 22, 28, and 34 are substantially the same as his arguments with respect to encoding claims 4, 10, and 16. Accordingly, it is my opinion that these claims are valid for the same reasons.

11. Dependent Claims 23, 29, and 35

129. I note that Dr. Freedman's opinions with respect to decoding claims 23, 29, and 35 are substantially the same as his arguments with respect to encoding claims 5, 11, and 17. Accordingly, it is my opinion that these claims are valid for the same reasons.

12. Dependent Claims 24, 30, and 36

130. I note that Dr. Freedman's opinions with respect to decoding claims 24, 30, and 36 are substantially the same as his arguments with respect to encoding claims 6, 12, and 18. Accordingly, it is my opinion that these claims are valid for the same reasons.

B. Ground 2 – Karczewicz-I and Karczewicz-II

131. I understand that Petitioners have asserted, and Dr. Freedman has opined, that the combination of Karczewicz-I and Karczewicz-II would render obvious the challenged claims of the '267 Patent. I disagree because a POSITA would not be motivated to modify the two references as Dr. Freedman describes.

1. Independent Claims 1, 7, and 13

a) A POSITA would not be motivated to combine and modify Karczewicz-I and Karczewicz-II as alleged.

132. Dr. Freedman's proposed motivation to combine and modify these references is fundamentally flawed. Nothing in Karczewicz-I or Karczewicz-II (or any other cited art) supports the idea of combining the two. Dr. Freedman's first argument for motivation to combine is that the Karczewicz references are both directed to video coding and recite "similar architectures." *See* Ex-1003 at ¶ 136. Dr. Freedman presents side-by-side figures from the two references, which I reproduce below:

Karczewicz-I

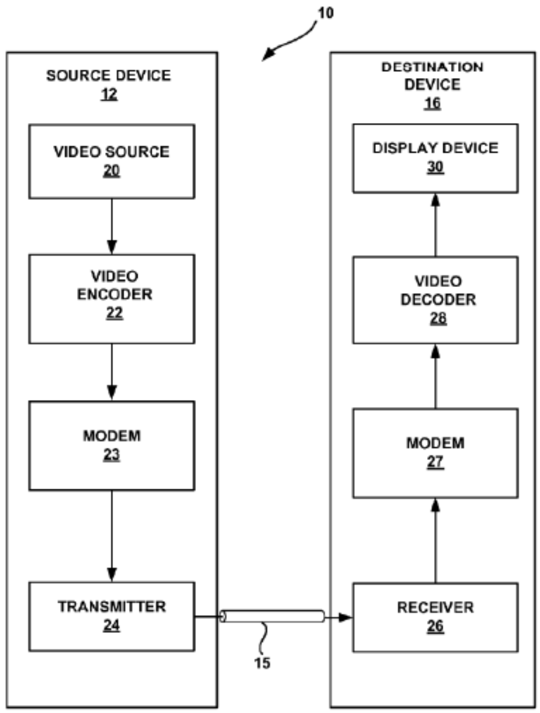


FIG. 1

Ex-1005, Fig. 1, ¶¶29-50.

Karczewicz-II

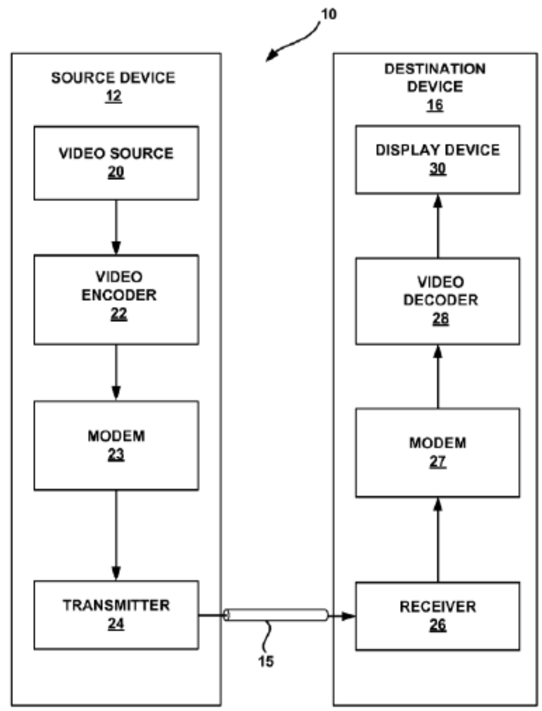


FIG. 1

Ex-1006, Fig. 1, ¶¶40-53.

133. The above figures represent the most abstracted view of a video transmission system that seems possible to me—one that applies to any video coding system which makes use of some means of telecommunication to transmit video between a "source" and "destination" device. Similarly, Dr. Freedman presents figures from both references depicting the same block diagram of a video coding system.

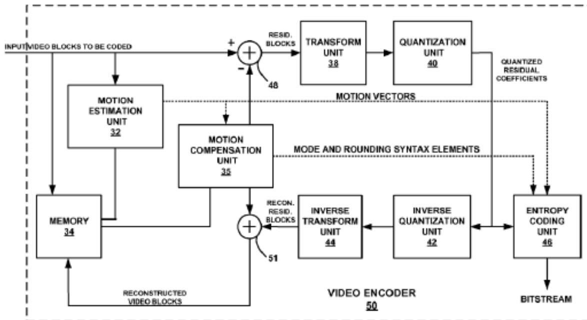


FIG. 2

Ex-1005, Fig. 2, ¶53 ("During the encoding process, video encoder 50 receives a video block to be coded, and motion estimation unit 32 and motion compensation unit 35 perform inter-predictive coding.").

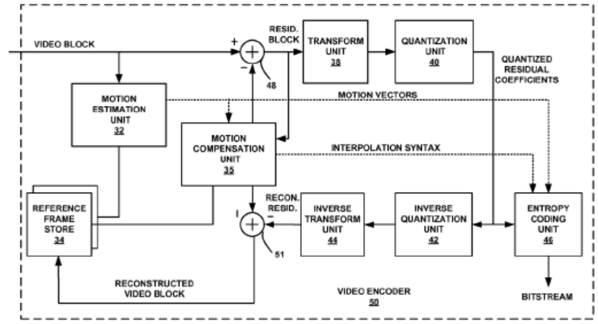


FIG. 2

Ex-1006, Fig. 2, ¶56 ("During the encoding process, video encoder 50 receives a video block to be coded, and motion estimation unit 32 and motion compensation unit 35 perform inter-predictive coding.").

134. These figures are, again, entirely generic, and depict basic components such as a "memory," "motion estimation unit," "motion compensation unit," and "entropy coding unit" which can be found in virtually any video coding system since the 1990s. Dr. Freedman agrees with me on this point, and describes Karczewicz-I's Figure 2 as "[t]he model of a typical general video encoder" that "has been used by major video encoding standards since the 1990s." Ex-1003 at ¶ 3. I note that both of the above figures disclose essentially the same set of basic components in the same configuration as those illustrated in my book (Ex-2012) at page 69, which I discuss above. Accordingly, that the same two figures are presented in both Karczewicz-I

and Karczewicz-II does not reflect their inherent suitability for combination so much as it reflects their sharing the same three inventors (Marta Karczewicz, Peisong Chen, and Yan Ye), and I would speculate, a desire to avoid unnecessary effort in producing introductory diagrams by those drafting the references. These figures would not provide any inspiration to a POSITA to combine and modify the two references. Rather, a POSITA would recognize the figures as depicting generic video coding systems that have existed since the 1990s. This consistent use of the same generic architecture over the course of decades would not lead a POSITA to search for a way to modify such a system.

135. The shared inventorship of the two Karczewicz references is also in my view highly relevant. During his deposition, Dr. Freedman denied any familiarity with the inventors, even saying that he could not be sure that they were persons of ordinary skill in the art. *See* Ex-2016 at 76:19-78:22. As a professional with years of experience in the video coding industry, however, I am familiar with all three inventors and indeed very familiar with Dr. Marta Karczewicz. Dr. Karczewicz is a well-known, world-leading expert in video compression, currently occupying the role of Vice President of Technology at Qualcomm, where she continues to make significant contributions to the field of video coding. Dr. Karczewicz has made many technical contributions to international standards, taking a very active role in working groups such as the Joint Video Team and Joint Coding Team – Video

Coding, as part of the MPEG and VCEG standardization committees. Dr. Karczewicz has been recognized by the European Patent Office as one of the "50 Leading Tech Voices" as part of the European Patent Office's celebration of 50 years of the European Patent Convention. Ex-2017. Though Dr. Freedman may not be sure if Dr. Karczewicz is a POSITA, in my opinion her skills, education, and experience would establish her as a person of ***extraordinary*** skill in the art. Accordingly, the idea that Dr. Karczewicz would have overlooked an "obvious" combination and modification of her own inventions, despite the supposed "architectural similarity" between them, lacks credibility.

136. Further, the disclosures of Karczewicz-II (which relate to subpixel interpolation—not bi-prediction) are coextensive with VCEG-AI33, a standards contribution co-authored by Dr. Karczewicz and corresponding to the same inventions. *See* Ex-1006 at ¶ 104, Table 7, 8 (showing the disclosed interpolation process and 16-bit registers); Ex-2001 at ¶¶ 1-4 (showing the same interpolation process and registers). VCEG-AI33 was "incorporated by reference in its entirety" into Karczewicz-I. *See* Ex-1005 at ¶ 37. Based on this, it is clear that Dr. Karczewicz had all the ingredients necessary to make the combination and modifications that Dr. Freedman proposes are obvious, and yet she and her co-authors did not do so. Therefore, I conclude that a POSITA with nothing more than a bachelor's degree and a few years of industry experience would not find the inventions of the '267

Patent obvious based on the disclosures of Karczewicz-I and Karczewicz-II. *See* Ex-1003 (stating that a POSITA would have "a (1) Bachelor's degree in electrical engineering, computer engineering, computer science, or a comparable field of study such as physics, and (2) approximately two to three years of practical experience with video encoding/decoding").

137. During deposition, Dr. Freedman speculated that the inventors of Karczewicz-I and Karczewicz-II may not have been aware that the applications had been filed in the first place. *See* Ex-2016 at 79:18-80:12 ("I am sorry to say that there are organizations that will file patent applications that may never have been reviewed or notified to their inventors . . . So, I cannot tell you whether those people actually even knew the patent application was being filed"). To the extent that Dr. Freedman intended to suggest that the reason that this supposedly "obvious" modification did not come about sooner was simply that the three co-authors did not realize the nature of their own work, I disagree. In my opinion, a more reasonable explanation is that Karczewicz-I and Karczewicz-II are not obvious to combine and modify in the way alleged by Dr. Freedman at all, and that it is only through the benefit of hindsight and reference to the inventions of the '267 Patent that Dr. Freedman draws the opposite conclusion.

138. As I discuss above, Karczewicz-I discloses improving conventional bi-prediction by adaptively selecting the offsets used during rounding performed when

calculating the final prediction and minimizing the number of offsets used. *See* Ex-1005 at ¶¶ 63-65. The reference explains that careful choice of rounding offset can improve bi-directional prediction. Thus Karczewicz-I proposes a potential solution to the problem of bidirectional prediction performance that does ***not*** involve increasing the precision of the two component predictions. Instead, Karczewicz-I adjusts the rounding offset to improve prediction performance but maintains all predictions at the original pixel precision and bit depth. Each of the predictions pred(i,j), pred0(i,j) and pred1(i,j) are retained at the original bit depth, as demonstrated by Karczewicz-I's calculation of pred(i,j). *See id.* at ¶¶ 64-66. This alone would teach away from the proposed combination and modifications because Karczewicz-I already discloses a solution to the problem of prediction performance and so a POSITA would not look to replace this or add to this solution in the way Dr. Freedman suggests.

139. Further, Karczewicz-II does not propose an improvement to the bi-prediction process itself. Instead, that reference discloses improving a different process, subpixel interpolation, by maintaining a higher precision for intermediate calculation of subpixel values. *See, e.g.,* Ex-1006 at ¶ 88.

140. Karczewicz-II does not disclose improving bi-prediction accuracy, at most it teaches how to improve the performance of interpolation. Karczewicz-II's process does not output higher-precision predictions, it outputs subpixel values at

the same 8-bit precision as the pixels used to obtain them. *See id.* at ¶ 102. Nothing in Karczewicz-I or Karczewicz-II states or hints that techniques of Karczewicz-I could be substituted for those of Karczewicz-II to yield predictable results for bi-prediction.

141. Further still, Dr. Freedman's description of how a POSITA would be motivated to modify Karczewicz-I by applying Karczewicz-II's "optimizations for averaging integer, half, and center pixels" are structured into three distinct scenarios, each of which operates differently and varies in the bit-depth of output prediction values. *See* Ex 1003 at ¶ 150. Rather than motivating a POSITA to modify Karczewicz-I as Dr. Freedman describes, these so-called scenarios (which are a subset of possible outcomes) present additional computational complexity in an encoder implementing Dr. Freedman's proposed system, and that would suggest to a POSITA that such a combination was undesirable.

142. During video coding, an encoder determines a motion vector describing the position of a block in a frame which closely matches the macroblock being coded. Ex-2012 at 31. This process involves searching previously coded frames for a block of pixels which, if used as the basis for a prediction, minimizes the chosen matching criteria—typically the residual formed when subtracting the prediction from the block being coded. *See id.* at 31-32. During bi-prediction, two such motion

vectors must be determined such that their combined predictions minimize the residual error, requiring further comparisons between potential best matches.

143. Dr. Freedman proposes that one of scenarios should apply based on the characteristics of both motion vectors used for bi-prediction. Ex-1003 at ¶ 150.

144. In his first scenario, corresponding to a half-pixel first motion vector and integer pixel second motion vector, he says that Karczewicz-II's Table 3 calculations should be applied, resulting in 13-bit prediction values. *Id.* at ¶¶ 151-153. In his second scenario, where the first motion vector points to a "center-pixel position" and the second motion vector points to a half-pixel position, he proposes using the same interpolation methodology to yield 15-bit values. *Id.* at ¶¶ 157-159. Finally, his third scenario applies to two half-pixel motion vectors and uses Karczewicz-II's Table 2 equations to produce a 15-bit prediction for each interpolated pixel. *Id.* at ¶¶ 163-165. Each of these scenarios results in a different value which may potentially have a different bit-depth.

145. As a result of these different potential interpolation equations, an encoder implementing Dr. Freedman's three scenarios would be forced to allocate additional space in which to store each possible intermediate prediction value computed for a first motion vector while the second prediction is calculated. Further, because an encoder must repeatedly test potential predictions to determine which minimizes residual, this storage cost would be magnified. In a software environment,

these costs would lead to additional memory and processing overhead which would be detrimental to the computational performance of an encoder. In a hardware environment, additional memory and logic to account for disparate scenarios translates to increased chip complexity and cost. Dr. Freedman does not consider or comment on these additional technical costs, noting that he was "not qualified to speak about the internals of any piece of hardware." Ex-2016 at 104:3-16. However, as I explained above, a POSITA would not choose to add computational complexity and additional storage requirements without a strong motivation to do so.

146. Because Dr. Freedman is unfamiliar with hardware, he instead discusses how such a system would be implemented at a software level, suggesting that it would be a "tiny change" that can be "decided with a switch statement" or an "'if' statement." *Id.* at 102:21-103:10. Dr. Freedman is suggesting adding "separate code branches" which would need to be written for bi-directional averaging, which the '267 Patent discusses as an undesirable feature of prior attempts at solving the rounding error problem. Ex-1001 at 4:21-25. Worse still, the '267 Patent criticizes the use of two branches, but Dr. Freedman's proposed modifications implicate three or more branches. A POSITA would be aware of the likely significant computational cost of adding further "code branches" to an operation (calculating bi-prediction

values) that may be carried out millions of times per second in a typical video coding application[8]. *See id.*

147. As I have discussed above, it is my opinion that a POSITA would not be motivated to make the modifications which Dr. Freedman proposes. In addition, I believe a POSITA would recognize that Dr. Freedman's proposal would incur considerable technical costs and disadvantages. Accordingly, it is my opinion that the challenged claims are not rendered obvious by the combination of Karczewicz-I and Karczewicz-II.

2. Dependent Claims 2, 8, and 14

148. Dependent claims 2, 8, and 14 depend on independent claims 1, 7, and 13 respectively. These claims include the additional limitation "wherein in an instance in which said first motion vector points to a subpixel, said first prediction is obtained by interpolation using pixel values of said first reference block." It is my opinion that a POSITA would not be motivated to modify Karczewicz-I and Karczewicz-II to make the inventions of claims 2, 8, and 14 for the same reasons as for the claims on which they depend.

[8] Consider a 30 frames per second High Definition video sequence. Each frame comprises 1920x1080 = 2073600 pixels. Assume (conservatively) that 20 frames per second use bi-prediction and that 10% of pixel values within these frames are predicted using bi-prediction. Over 4 million pixel values per second are predicted using bi-prediction at the decoder.

3. Dependent Claims 3, 9, and 15

149. Dependent claims 3, 9, and 15 depend on claims 2, 8, and 14 respectively. These claims include the additional limitation "wherein said first prediction is obtained by interpolation using values of said first reference block by: right shifting a sum of a P-tap filter using values of said first reference block." It is my opinion that a POSITA would not be motivated to modify Karczewicz-I and Karczewicz-II to make the inventions of claims 3, 9, and 15 for the same reasons as for the claims on which they depend.

4. Dependent Claims 4, 10, and 16

150. Dependent claims 4, 10, and 16 depend on claims 2, 8, and 14 respectively. These claims include the additional limitation "wherein in an instance in which said second motion vector points to an integer sample, said second prediction is obtained by shifting values of said second reference block to the left." It is my opinion that a POSITA would not be motivated to modify Karczewicz-I and Karczewicz-II to make the inventions of claims 4, 10, and 16 for the same reasons as for the claims on which they depend.

5. Dependent Claims 5, 11, and 17

151. Dependent claims 5, 11, and 17 depend on claims 1, 7, and 13 respectively. These claims include the additional limitation of decreasing the precision of the combined prediction by shifting bits of the combined prediction to

the right by inserting a rounding offset to the combined prediction before said decreasing. It is my opinion that a POSITA would not be motivated to modify Karczewicz-I and Karczewicz-II to make the inventions of claims 5, 11, and 17 for the same reasons as for the claims on which they depend.

6. Dependent Claims 6, 12, and 18

152. Dependent claims 6, 12, and 18 depend on claims 1, 7, and 13 respectively. These claims include the additional limitation "wherein the first precision indicates a number of bits needed to represent the values of the pixels, and the second precision indicates the number of bits needed to represent values of said first prediction and values of said second prediction." It is my opinion that a POSITA would not be motivated to modify Karczewicz-I and Karczewicz-II to make the inventions of claims 6, 12, and 18 for the same reasons as for the claims on which they depend.

7. Independent Claims 19, 25, and 31

153. Independent claims 19, 25, and 31 are decoding claims corresponding to encoding claims 1, 7, and 13. I understand that Dr. Freedman's opinions with respect to these claims are largely the same as the ones he offers for their encoding counterparts, and so I disagree with him for the same reasons.

154. Additionally, I note that the same technical costs associated with Dr. Freedman's three proposed scenarios for modifying Karczewicz-I using

Karczewicz-II are also applicable to a decoder. A decoder which generates predictions using Dr. Freedman's three scenarios would necessitate that bitstreams be conducted by an encoder using the same logic to generate its predictions and determine motion vectors to be encoded. A POSITA would understand that creating a decoder implementing Dr. Freedman's proposed modifications would mean incurring the same encoding costs that I described with respect to claims 1, 7, and 13. Accordingly, it is my opinion that a POSITA would not be motivated modify Karczewicz-I and Karczewicz-II as Dr. Freedman states.

8. Dependent Claims 20, 26, and 32

155. I note that Dr. Freedman's opinions with respect to decoding claims 20, 26, and 32 are substantially the same as his arguments with respect to encoding claims 2, 8, and 14. Accordingly, it is my opinion that these claims are valid for the same reasons.

9. Dependent Claims 21, 27, and 33

156. I note that Dr. Freedman's opinions with respect to decoding claims 21, 27, and 33 are substantially the same as his arguments with respect to encoding claims 3, 9, and 15. Accordingly, it is my opinion that these claims are valid for the same reasons.

10. Dependent Claims 22, 28, and 34

157. I note that Dr. Freedman's opinions with respect to decoding claims 22, 28, and 34 are substantially the same as his arguments with respect to encoding claims 4, 10, and 16. Accordingly, it is my opinion that these claims are valid for the same reasons.

11. Dependent Claims 23, 29, and 35

158. I note that Dr. Freedman's opinions with respect to decoding claims 23, 29, and 35 are substantially the same as his arguments with respect to encoding claims 5, 11, and 17. Accordingly, it is my opinion that these claims are valid for the same reasons.

12. Dependent Claims 24, 30, and 36

159. I note that Dr. Freedman's opinions with respect to decoding claims 24, 30, and 36 are substantially the same as his arguments with respect to encoding claims 6, 12, and 18. Accordingly, it is my opinion that these claims are valid for the same reasons.

VIII. Conclusion

160. This Declaration represents the opinions that I have formed to date based on the material currently available to me. I reserve the right to revise, supplement, and/or amend my opinions stated in this Declaration based on new information that becomes available to me and on my continuing analysis of the materials available.

161. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that these statements are made with the knowledge that willful false statements and the like so made are punishable by fine, imprisonment, or both, under 18 U.S.C. § 1001.

Dated: December 16, 2024

Iain E. Richardson

Dr. Iain Richardson