

THE **H.264**  
ADVANCED VIDEO  
COMPRESSION STANDARD

SECOND EDITION

IAIN E. RICHARDSON

 WILEY

# **Contents**

**[About the Author](#)**

**[Preface](#)**

**[Glossary](#)**

**[List of Figures](#)**

**[List of Tables](#)**

**[1 Introduction](#)**

**[1.1 A change of scene](#)**

**[1.2 Driving the change](#)**

**[1.3 The role of standards](#)**

**[1.4 Why H.264 Advanced Video Coding is important](#)**

**[1.5 About this book](#)**

**[1.6 Reference](#)**

**[2 Video formats and quality](#)**

**[2.1 Introduction](#)**

**[2.2 Natural video scenes](#)**

**[2.3 Capture](#)**

**[2.4 Colour spaces](#)**

**[2.5 Video formats](#)**

**[2.6 Quality](#)**

**[2.7 Summary](#)**

## **2.8 References**

# **3 Video coding concepts**

## **3.1 Introduction**

## **3.2 Video CODEC**

## **3.3 Prediction model**

## **3.4 Image model**

## **3.5 Entropy coder**

## **3.6 The hybrid DPCM/DCT video CODEC model**

## **3.7 Summary**

## **3.8 References**

# **4 What is H.264?**

## **4.1 Introduction**

## **4.2 What is H.264?**

## **4.3 How does an H.264 codec work?**

## **4.4 The H.264/AVC Standard**

## **4.5 H.264 Profiles and Levels**

## **4.6 The H.264 Syntax**

## **4.7 H.264 in practice**

## **4.8 Summary**

## **4.9 References**

# **5 H.264 syntax**

## **5.1 Introduction**

## **5.2 H.264 syntax**

## **5.3 Frames, fields and pictures**

## **5.4 NAL unit**

- [\*\*5.5 Parameter Sets\*\*](#)
- [\*\*5.6 Slice layer\*\*](#)
- [\*\*5.7 Macroblock layer\*\*](#)
- [\*\*5.8 Summary\*\*](#)
- [\*\*5.9 References\*\*](#)

## [\*\*6 H.264 Prediction\*\*](#)

- [\*\*6.1 Introduction\*\*](#)
- [\*\*6.2 Macroblock prediction\*\*](#)
- [\*\*6.3 Intra prediction\*\*](#)
- [\*\*6.4 Inter prediction\*\*](#)
- [\*\*6.5 Loop filter\*\*](#)
- [\*\*6.6 Summary\*\*](#)
- [\*\*6.7 References\*\*](#)

## [\*\*7 H.264 transform and coding\*\*](#)

- [\*\*7.1 Introduction\*\*](#)
- [\*\*7.2 Transform and quantization\*\*](#)
- [\*\*7.3 Block scan orders\*\*](#)
- [\*\*7.4 Coding\*\*](#)
- [\*\*7.5 Summary\*\*](#)
- [\*\*7.6 References\*\*](#)

## [\*\*8 H.264 conformance, transport and licensing\*\*](#)

- [\*\*8.1 Introduction\*\*](#)
- [\*\*8.2 Conforming to the Standard\*\*](#)
- [\*\*8.3 H.264 coding tools for transport support\*\*](#)



- [\*\*8.4 Transport of H.264 data\*\*](#)
- [\*\*8.5 Supplemental Information\*\*](#)
- [\*\*8.6 Licensing H.264/AVC\*\*](#)
- [\*\*8.7 Summary\*\*](#)
- [\*\*8.8 References\*\*](#)

## [\*\*9 H.264 performance\*\*](#)

- [\*\*9.1 Introduction\*\*](#)
- [\*\*9.2 Experimenting with H.264\*\*](#)
- [\*\*9.3 Performance comparisons\*\*](#)
- [\*\*9.4 Rate control\*\*](#)
- [\*\*9.5 Mode selection\*\*](#)
- [\*\*9.6 Low complexity coding\*\*](#)
- [\*\*9.7 Summary\*\*](#)

## [\*\*10 Extensions and directions\*\*](#)

- [\*\*10.1 Introduction\*\*](#)
- [\*\*10.2 Scalable Video Coding\*\*](#)
- [\*\*10.3 Multiview Video Coding\*\*](#)
- [\*\*10.4 Configurable Video Coding\*\*](#)
- [\*\*10.5 Beyond H.264/AVC\*\*](#)
- [\*\*10.6 Summary\*\*](#)
- [\*\*10.7 References\*\*](#)

## [\*\*Index\*\*](#)

# **THE H.264 ADVANCED VIDEO COMPRESSION STANDARD**

**Second Edition**

**Iain E. Richardson**

*Vcodex Limited, UK*



A John Wiley and Sons, Ltd., Publication

This edition first published 2010

© 2010, John Wiley & Sons, Ltd

First Edition published in 2003

*Registered office*

John Wiley & Sons Ltd, The Atrium, Southern Gate,  
Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at [www.wiley.com](http://www.wiley.com).

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or

other expert assistance is required, the services of a competent professional should be sought.

*Library of Congress Cataloguing-in-Publication Data*

Richardson, Iain E. G.

The H.264 advanced video compression standard/Iain E. Richardson. - 2nd ed.

p. cm.

Rev. ed. of: H.264 and MPEG-4 video compression. c2003.

Includes bibliographical references and index.

ISBN 978-0-470-51692-8 (pbk.)

1. Digital video-Standards. 2. Video compression-Standards.  
3. MPEG (Video coding standard) 4. Multimedia systems. I.

Richardson, Iain E. G. H.264 and MPEG-4 video compression.

II. Title.

TK6680.5.R52 2010

006.6'96-dc22

2009054387

ISBN: 978-0-470-51692-8

*To Pat*

*Language is living, but what is most important goes deeper  
than words.*



## ***About the Author***

Professor Iain Richardson is an internationally known expert on the MPEG and H.264 video compression standards.

The author of *H.264 and MPEG-4 Video Compression*, a widely cited work in the research literature, Professor Richardson has written two further books and over 70 journal and conference papers on video compression. He regularly advises companies on video compression technology, video coding patents and company acquisitions in the video coding industry. Professor Richardson leads an internationally renowned video coding research team, contributes to the MPEG industry standards group and is sought after as an expert witness. Based in Aberdeen, Scotland, he regularly travels to the US and Europe.

# ***Preface***

The last decade has seen a quiet revolution in digital video technology. Digital video is everywhere: on our televisions, our DVD and Blu-Ray players, our computers, our music players and our mobile handsets. Only recently, a video image in a web page was an unusual sight. Nowadays, many of us are just as likely to catch the latest news on the web as on the TV. With the explosion of digital video applications, a billion-dollar industry has developed and expanded, with new companies and niche markets emerging, thriving and disappearing faster than anyone can easily track. Video compression is essential to all of these applications and markets, and the H.264 format is considered by many to be the state of the art in video compression.

When I wrote the first edition of this book in 2003, H.264 Advanced Video Compression had just been published as an International Standard and it was hard to predict its impact on industry. Its predecessor, MPEG-4 Visual, had arguably failed to live up to its promise, with only limited adoption in the market. Since 2003, the significant performance improvements that are built into H.264 have made it the clear successor to the older MPEG video standards in many applications, from mobile video to High Definition broadcasting. At the time of writing, the MPEG and VCEG standards committees are debating the possible successor to H.264. It is likely to be several years before a new standard is released, and several years after that before H.264 begins to become obsolete.

This book is intended to be a practical, accessible and unbiased guide to the H.264 video compression standard. As always, I have chosen to explain the details of H.264 in my own way, concentrating on what I feel is important to the engineer, researcher or student who needs a 'way in' to this complex yet important technical subject. This book is

not the final word on H.264. By definition, that final word is provided by the standard itself and I advise any serious developer or implementer of H.264 to get hold of a copy of the standard. There is a need for a guidebook to the standard that explains the concepts, tools, benefits and disadvantages of the format, just as a good guidebook helps the tourist to get to know a foreign country and to become more at home there. Some visitors may be disappointed that their favourite subject is not covered in as much depth as they would like. I have made a deliberate choice to cover certain topics such as Scalable and Multiview Video Coding only briefly as they are still, in my view, in the early stages of practical implementation.

My sincere thanks to the many people who have helped to shape this book, including the readers of my earlier books who told me what they liked and what they wanted; the many companies and individuals who have asked me to solve their video compression problems; Kourosh Soroushian for discussions on Hypothetical Reference Decoders; Abharana Bhat, MajaBystrom, SamJansen, Sampath Kannangara and Yafan Zhao for reading and commenting on draft chapters; Gary Sullivan for many comments, corrections, suggestions and discussions; Nicky, Simone and the editorial team at John Wiley & Sons; and to Pat for reading the manuscript, cracking the whip and making me finish it.

I hope that you find the book useful; more importantly, I hope you enjoy it. Visit my website at [www.vcodex.com](http://www.vcodex.com) and *tell me what you think*.

Iain Richardson  
Aberdeen, 2010

# Glossary

4:2:0 (sampling)	Sampling method: chrominance components have half the horizontal and vertical resolution of luminance component
4:2:2 (sampling)	Sampling method: chrominance components have half the horizontal resolution of luminance component
4:4:4 (sampling)	Sampling method: chrominance components have same resolution as luminance component
access unit	Complete coded frame or field
arithmetic coding	Coding method to reduce redundancy
artefact	Visual distortion in an image
ASO	Arbitrary Slice Order, in which slices may be coded out of raster sequence
block	Region of macroblock
block matching	Motion estimation carried out on rectangular picture areas
blocking	Square or rectangular distortion areas in an image
B slice	Coded slice predicted using bidirectional motion compensation
CABAC	Context-based Adaptive Binary Arithmetic Coding
CAVLC	Context Adaptive Variable Length Coding
chrominance or chroma	Colour difference component
CIF	Common Intermediate Format, a colour image format
CODEC	COder/DECoder pair
Coded Picture Buffer (CPB)	Buffer containing coded frames or fields
colour space	Method of representing colour images
DCT	Discrete Cosine Transform, a mathematical transform and/or its practical approximation(s)
direct prediction	A coding mode in which no motion vector is transmitted
DPCM	Differential Pulse Code Modulation
DSCQS	Double Stimulus Continuous Quality Scale, a scale and method for subjective quality measurement
DWT	Discrete Wavelet Transform
entropy coding	Coding method to reduce redundancy

error concealment	Post-processing of a decoded image to remove or reduce visible error effects
Exp-Golomb or ExpG	Exponential Golomb variable length codes
Field	Odd- or even-numbered lines from an interlaced video sequence
FMO	Flexible Macroblock Order, in which macroblocks may be coded out of raster sequence
Full Search	A motion estimation algorithm
Fully Configurable Video Coding	A framework for video coding in which a codec may be completely re-configured during a communication session
GOP	Group of Pictures, a set of coded video images
H.261	A video coding standard
H.263	A video coding standard
H.264	A video coding standard
HDTV	High Definition Television
Huffman coding	Coding method to reduce redundancy
HVS	Human Visual System, the system by which humans perceive and interpret visual images
hybrid (CODEC)	CODEC model featuring motion compensation and transform
Hypothetical Reference Decoder (HRD)	Decoder 'model' that may be used to test bitstream conformance
IEC	International Electrotechnical Commission, a standards body
inter (coding)	Coding of video frames using temporal prediction or compensation
interlaced (video)	Video data represented as a series of fields
intra (coding)	Coding of video frames without temporal prediction
I slice	Slice coded without reference to any other frame
ISO	International Standards Organisation, a standards body
ITU	International Telecommunication Union, a standards body
JPEG	Joint Photographic Experts Group, a committee of ISO (also an image coding standard)
latency	Delay through a communication system
Level	A set of conformance parameters (applied to a Profile)
loop filter	Spatial filter placed within encoding or decoding feedback



	loop
luminance or luma	Monochrome or brightness component
Macroblock	Region of frame coded as a unit (usually 16 × 16 pixels in the original frame)
Macroblock partition	Region of macroblock with its own motion vector
Macroblock sub-partition	Region of macroblock with its own motion vector
motion compensation	Prediction of a video frame with modelling of motion
motion estimation	Estimation of relative motion between two or more video frames
motion vector	Vector indicating a displaced block or region to be used for motion compensation
MPEG	Motion Picture Experts Group, a committee of ISO/IEC
MPEG-1	A multimedia coding standard
MPEG-2	A multimedia coding standard
MPEG-4	A multimedia coding standard
MVC	Multiview Video Coding, in which multiple views of a scene may be jointly coded
NAL	Network Abstraction Layer
objective quality	Visual quality measured by algorithm(s)
Picture (coded)	Coded (compressed) video frame
P-picture (slice)	Coded picture (or slice) using motion-compensated prediction from one reference frame
Profile	A set of functional capabilities (of a video CODEC)
progressive (video)	Video data represented as a series of complete frames
PSNR	Peak Signal to Noise Ratio, an objective quality measure
QCIF	Quarter Common Intermediate Format
quantize	Reduce the precision of a scalar or vector quantity
rate control	Control of bit rate of encoded video signal
rate-distortion	Measure of CODEC performance (distortion at a range of coded bit rates)
RBSP	Raw Byte Sequence Payload
RVC	Reconfigurable Video Coding, a framework for video coding in which a decoder may be constructed from pre-defined

	Functional Units.
RGB	Red/Green/Blue colour space
ringing (artefacts)	'Ripple'-like artefacts around sharp edges in a decoded image
RTP	Real Time Protocol, a transport protocol for real-time data
scalable coding	Coding a signal into a number of layers
SVC	Scalable Video Coding
SI slice	Intra-coded slice used for switching between coded bitstreams (H.264)
Slice	A region of a coded picture
SP slice	Inter-coded slice used for switching between coded bitstreams
statistical redundancy	Redundancy due to the statistical distribution of data
studio quality	Lossless or near-lossless video quality
subjective quality	Visual quality as perceived by human observer(s)
subjective redundancy	Redundancy due to components of the data that are subjectively insignificant
sub-pixel (motion compensation)	Motion-compensated prediction from a reference area that may be formed by interpolating between integer-valued pixel positions
test model	A software model and document that describe a reference implementation of a video coding standard
texture	Image or residual data
tree-structured motion compensation	Motion compensation featuring a flexible hierarchy of partition sizes
VCEG	Video Coding Experts Group, a committee of ITU
VCL	Video Coding Layer
video packet	Coded unit suitable for packetization
VLC	Variable Length Code
VLD	Variable Length Decoder
VLE	Variable Length Encoder
VLSI	Very Large Scale Integrated circuit
VQEG	Video Quality Experts Group
weighted prediction	Motion compensation in which the prediction samples from two references are scaled

YCrCb

Luminance/Red chrominance/Blue chrominance colour space

# ***List of Figures***

- [1.1](#) Video coding scenarios, one-way
- [1.2](#) Video coding scenario, two-way
- [2.1](#) Still image from natural video scene
- [2.2](#) Spatial and temporal sampling of a video sequence
- [2.3](#) Image with two sampling grids
- [2.4](#) Image sampled at coarse resolution (black sampling grid)
- [2.5](#) Image sampled at finer resolution (grey sampling grid)
- [2.6](#) Interlaced video sequence
- [2.7](#) Top field
- [2.8](#) Bottom field
- [2.9](#) Red, Green and Blue components of colour image
- [2.10](#) Cr, Cg and Cb components
- [2.11](#) 4:2:0, 4:2:2 and 4:4:4 sampling patterns (progressive)
- [2.12](#) Allocation of 4:2:0 samples to top and bottom fields
- [2.13](#) Video frame sampled at range of resolutions
- [2.14](#) SD and HD formats
- [2.15](#) DSCQS testing system
- [2.16](#) PSNR examples: (a) Original; (b) 30.6dB; (c) 28.3dB
- [2.17](#) Image with blurred background (PSNR = 27.7dB)
- [3.1](#) Encoder/Decoder
- [3.2](#) Spatial and temporal correlation in a video sequence
- [3.3](#) Video encoder block diagram
- [3.4](#) Frame 1
- [3.5](#) Frame 2
- [3.6](#) Difference
- [3.7](#) Optical flow
- [3.8](#) Macroblock (4:2:0)
- [3.9](#) Motion estimation
- [3.10](#) Frame 1
- [3.11](#) Frame
- [3.12](#) Residual: no motion compensation
- [3.13](#) Residual: 16 × 16 block size
- [3.14](#) Residual: 8 × 8 block size
- [3.15](#) Residual: 4 × 4 block size

- [3.16](#) Close-up of reference region
- [3.17](#) Reference region interpolated to half-pixel positions
- [3.18](#) Integer, half-pixel and quarter-pixel motion estimation
- [3.19](#) Residual:  $4 \times 4$  blocks, 1/2-pixel compensation
- [3.20](#) Residual:  $4 \times 4$  blocks, 1/4-pixel compensation
- [3.21](#) Motion vector map:  $16 \times 16$  blocks, integer vectors
- [3.22](#) Motion vector map:  $4 \times 4$  blocks, 1/4-pixel vectors
- [3.23](#) Intra prediction: available samples
- [3.24](#) Intra prediction: spatial extrapolation
- [3.25](#) 2D autocorrelation function of image
- [3.26](#) 2D autocorrelation function of residual
- [3.27](#) Spatial prediction (DPCM)
- [3.28](#)  $4 \times 4$  DCT basis patterns
- [3.29](#)  $8 \times 8$  DCT basis patterns
- [3.30](#) Image section showing  $4 \times 4$  block
- [3.31](#) Close-up of  $4 \times 4$  block; DCT coefficients
- [3.32](#) Block reconstructed from (a) 1, (b) 2, (c) 3, (d) 5 coefficients
- [3.33](#) Two-dimensional wavelet decomposition process
- [3.34](#) Image after one level of decomposition
- [3.35](#) Two-stage wavelet decomposition of image
- [3.36](#) Five-stage wavelet decomposition of image
- [3.37](#) Scalar quantizers: linear; non-linear with dead zone
- [3.38](#) Vector quantization
- [3.39](#)  $8 \times 8$  DCT coefficient distribution (frame)
- [3.40](#) Residual field picture
- [3.41](#)  $8 \times 8$  DCT coefficient distribution (field)
- [3.42](#) Zigzag scan example: frame block
- [3.43](#) Zigzag scan example: field block
- [3.44](#) Wavelet coefficient and 'children'
- [3.45](#) Motion vector prediction candidates
- [3.46](#) Generating the Huffman code tree: Sequence 1 motion vectors
- [3.47](#) Huffman tree for sequence 2 motion vectors
- [3.48](#) MPEG4 TCOEF VLCs (partial)
- [3.49](#) Sub-range example
- [3.50](#) Arithmetic coding example
- [3.51](#) DPCM/DCT video encoder



- [3.52](#) DPCM/DCT video decoder
- [3.53](#) Input frame  $F_n$
- [3.54](#) Reconstructed reference frame  $F'_{n-1}$
- [3.55](#) Residual  $F_n - F'_{n-1}$ : no motion compensation
- [3.56](#)  $16 \times 16$  motion vectors superimposed on frame
- [3.57](#) Motion compensated reference frame
- [3.58](#) Motion compensated residual frame
- [3.59](#) Original macroblock: luminance
- [3.60](#) Residual macroblock: luminance
- [3.61](#) DCT coefficient magnitudes: top-right  $8 \times 8$  block
- [3.62](#) Comparison of original and decoded residual blocks
- [3.63](#) Decoded frame  $F'_n$
- [4.1](#) The H.264 video coding and decoding process
- [4.2](#) Video coding: source frames, encoded bitstream, decoded frames
- [4.3](#) Video codec: high level view
- [4.4](#) Typical H.264 encoder
- [4.5](#) Typical H.264 decoder
- [4.6](#) Prediction: flow diagram
- [4.7](#) Intra prediction
- [4.8](#) Original macroblock, intra prediction and residual
- [4.9](#) Inter prediction
- [4.10](#) Original macroblock, inter prediction and residual
- [4.11](#) Forward transform
- [4.12](#) Quantization example
- [4.13](#) Example: Block, transform coefficients, quantized coefficients
- [4.14](#) Rescaling example
- [4.15](#) Inverse transform: combining weighted basis patterns to create a  $4 \times$  image block
- [4.16](#) Rescaled coefficients and inverse transform output
- [4.17](#) Reconstruction flow diagram
- [4.18](#) Profiles and Levels: example
- [4.19](#) H.264 syntax: overview
- [4.20](#) Syntax example: P-macroblock
- [4.21](#) P-macroblock decoding process
- [4.22](#) Residual luma and chroma coefficient blocks
- [4.23](#) A video frame compressed at the same bitrate using MPEG-2 (left),

MPEG-4 Visual (centre) and H.264 compression (right)

- [5.1](#) Syntax overview
- [5.2](#) Picture handling in H.264, overview
- [5.3](#) Decoded Picture Buffer and picture orders
- [5.4](#) Display order: Type 0 example
- [5.5](#) Display order: Type 1 example
- [5.6](#) List 0 and List 1 ordering: example
- [5.7](#) Default reference pictures: example
- [5.8](#) Reference picture re-ordering syntax, simplified overview
- [5.9](#) Picture Adaptive Frame Field Coding example
- [5.10](#) Frame with macroblock pairs
- [5.11](#) MB pair coded using Macroblock Adaptive Frame Field Coding
- [5.12](#) Example: Sequence and Picture Parameter Sets
- [5.13](#) Macroblock layer syntax overview
- [5.14](#) mb pred and sub mb pred syntax overview
- [5.15](#) Residual data syntax overview
- [5.16](#) Block scanning order,  $4 \times 4$  transform, 4:2:0 sampling
- [5.17](#) Block scanning order,  $8 \times 8$  luma transform, 4:2:0 sampling
- [5.18](#) Residual CAVLC block syntax overview
- [5.19](#) Residual CABAC block syntax overview
- [5.20](#) P macroblock, example 2
- [5.21](#) P macroblock example 4
- [5.22](#) B macroblock, example 5
- [6.1](#) Example of macroblock types and prediction sources
- [6.2](#) Intra prediction: adjacent blocks example
- [6.3](#) Intra prediction source samples,  $4 \times 4$  or  $8 \times 8$  luma blocks
- [6.4](#) Intra prediction source samples, chroma or  $16 \times 16$  luma blocks
- [6.5](#) Example of intra block size choices, CIF, Baseline Profile. Reproduced by permission of Elecard.
- [6.6](#) QCIF frame with highlighted macroblock
- [6.7](#) Predicted luma frame formed using H.264 intra prediction
- [6.8](#) Residual after subtracting intra prediction
- [6.9](#)  $4 \times 4$  luma block to be predicted
- [6.10](#) Labelling of prediction samples,  $4 \times 4$  prediction
- [6.11](#)  $4 \times 4$  intra prediction modes
- [6.12](#) Prediction blocks,  $4 \times 4$  modes 0-8

- [6.13](#) Intra 16 × 16 prediction modes
- [6.14](#) 16 × 16 macroblock
- [6.15](#) Prediction blocks, intra 16 × 16 modes 0–3
- [6.16](#) Intra mode prediction example
- [6.17](#) P macroblock prediction example
- [6.18](#) Example of integer and sub-pixel prediction
- [6.19](#) Current region
- [6.20](#) 4 × 4 block to be predicted
- [6.21](#) Reference region
- [6.22](#) Prediction from integer samples
- [6.23](#) Reference region, half-pixel interpolated
- [6.24](#) Prediction from interpolated samples
- [6.25](#) Interpolation of luma half-pel positions
- [6.26](#) Interpolation of luma quarter-pel positions
- [6.27](#) Luma region interpolated to quarter-pel positions
- [6.28](#) Interpolation of chroma eighth-pel positions
- [6.29](#) Macroblock partitions and sub-macroblock partitions
- [6.30](#) Current and neighbouring partitions: same partition sizes
- [6.31](#) Current and neighbouring partitions: different partition sizes
- [6.32](#) Scaled motion vector example
- [6.33](#) Temporal direct motion vector example
- [6.34](#) Forming a motion compensated prediction
- [6.35](#) Biprediction example
- [6.36](#) MBAFF: prediction from corresponding field
- [6.37](#) P slice showing partition choices. Reproduced by permission of Elecard
- [6.38](#) B slice showing macroblock modes. Light-shaded circles are skipped macroblocks. Reproduced by permission of Elecard.
- [6.39](#) Inter prediction example, P slice
- [6.40](#) Inter prediction example, B slice
- [6.41](#) Low delay prediction structure
- [6.42](#) ‘Classic’ Group of Pictures prediction structure
- [6.43](#) IPPP... with multiple reference pictures
- [6.44](#) Hierarchical GOP structure
- [6.45](#) Edge filtering order in a macroblock
- [6.46](#) Pixels adjacent to vertical and horizontal boundaries
- [6.47](#) 16 × 16 luma macroblock showing block edges

- [6.48](#) Original frame, 'violin' frame 2
- [6.49](#) Reconstructed, QP=36, no filter
- [6.50](#) Reconstructed, QP=36, with filter
- [6.51](#) Reconstructed, QP=32, no filter
- [6.52](#) Reconstructed, QP=32, with filter
- [7.1](#) Re-scaling and inverse transform
- [7.2](#) Forward transform and quantization
- [7.3](#) Luma forward transform: default
- [7.4](#) Luma inverse transform: default
- [7.5](#) Luma forward transform: Intra 16 × 16 mode
- [7.6](#) Luma inverse transform: Intra 16 × 16 mode
- [7.7](#) Luma forward transform: 8 × 8 transform
- [7.8](#) Luma inverse transform: 8 × 8 transform
- [7.9](#) Chroma forward transform: 4:2:0 macroblock
- [7.10](#) Chroma inverse transform: 4:2:0 macroblock
- [7.11](#) Chroma forward transform: 4:2:2 macroblock
- [7.12](#) Chroma inverse transform: 4:2:2 macroblock
- [7.13](#) Development of the forward transform and quantization process
- [7.14](#) Development of the rescaling and inverse transform process
- [7.15](#) Quantization parameter QP vs. effective quantizer step size, logarithmic y-axis
- [7.16](#) Frequency dependent quantization, 4 × 4 block
- [7.17](#) Progressive scan orders for 4 × 4 and 8 × 8 blocks
- [7.18](#) Field scan orders for 4 × 4 and 8 × 8 blocks
- [7.19](#) CAVLC encoder overview
- [7.20](#) CABAC coding process overview
- [8.1](#) Baseline, Constrained Baseline, Extended and Main Profiles
- [8.2](#) Main and High Profiles
- [8.3](#) Main and Intra Profiles
- [8.4](#) Selected Level constraints
- [8.5](#) Selected display resolutions
- [8.6](#) H.264 encoder and decoder buffers
- [8.7](#) Hypothetical Reference Decoder (HRD)
- [8.8](#) HRD example 1: encoder buffer
- [8.9](#) HRD example 1: decoder CPB
- [8.10](#) HRD Example 2: encoder buffer

- [8.11](#) HRD Example 2: decoder CPB
- [8.12](#) HRD Example 3: encoder buffer
- [8.13](#) HRD Example 3: decoder CPB
- [8.14](#) Bitstream conformance testing
- [8.15](#) Decoder conformance testing
- [8.16](#) Arbitrary Slice Order: Example
- [8.17](#) FMO: Interleaved map, QCIF, 3 slice groups
- [8.18](#) FMO: Dispersed macroblock map, QCIF, 4 slice groups
- [8.19](#) FMO: Foreground and Background map, 4 slice groups
- [8.20](#) FMO: Box-out, Raster and Wipe maps
- [8.21](#) Switching streams using I-slices
- [8.22](#) Switching streams using SP-slices
- [8.23](#) Encoding SP-slice A2 (simplified)
- [8.24](#) Encoding SP-slice B2 (simplified)
- [8.25](#) Decoding SP-slice A2 (simplified)
- [8.26](#) Encoding SP-slice AB2 (simplified)
- [8.27](#) Decoding SP-slice AB2
- [8.28](#) Fast-forward using SP-slices
- [8.29](#) Encapsulation of H.264 syntax elements
- [8.30](#) MPEG-2 Transport Stream
- [8.31](#) RTP packet structure (simplified)
- [8.32](#) ISO Media File
- [8.33](#) Block diagram from US patent 3679821 (redrawn)
- [8.34](#) Issued US patents including the terms 'video coding' or 'video compression', 1990–2007. Source: USPTO patent database.
- [9.1](#) JM software operation
- [9.2](#) Planar YCbCr file format, 4:2:0 sampling
- [9.3](#) JM encoder configuration file
- [9.4](#) JM encoder output display
- [9.5](#) Original, reconstructed and decoded frames, container.qcif, QP=32
- [9.6](#) Section of coded frame, JM encoder (left), x264 encoder (right)
- [9.7](#) Screenshot of an H.264 stream analyzer: Baseline Profile frame. Reproduced by permission of Elecard
- [9.8](#) Screenshot of stream analyzer: Main Profile frame. Courtesy of Elecard
- [9.9](#) Foreman/QCIF/Basic complexity
- [9.10](#) Foreman/QCIF/Basic complexity and options

- [9.11](#) Foreman/QCIF/Basic and Medium Complexity
- [9.12](#) Foreman/QCIF/Medium complexity with rate control
- [9.13](#) Foreman, QCIF sequences: coding time
- [9.14](#) QCIF sequences: rate vs. PSNR
- [9.15](#) CIF sequences: rate vs. PSNR
- [9.16](#) CIF sequences: coding time
- [9.17](#) Rate-distortion comparison of prediction structures
- [9.18](#) Sample frames from sequences using different prediction structures, coded at 280kbps
- [9.19](#) Carphone, QCIF: H.264 vs. MPEG-4 Visual
- [9.20](#) Frame from 'Foreman' sequence showing macroblock sizes
- [9.21](#) Encoder with rate feedback
- [9.22](#) Bitrate allocation for rate control
- [9.23](#) Foreman, QCIF, 100 frames: coded bitrate
- [9.24](#) Foreman, QCIF, 100 frames: QP per frame
- [9.25](#) Foreman, QCIF, 100 frames: Luma PSNR per frame
- [9.26](#) Available macroblock prediction modes
- [9.27](#) Rate and MSE costs for different coding options
- [9.28](#) Rate, quality and complexity
- [10.1](#) Multiple streams/simulcast
- [10.2](#) Multiple streams/scalable
- [10.3](#) Overview of scalability types
- [10.4](#) Temporally scalable sequence, 3 layers
- [10.5](#) Decoding a temporally scalable sequence
- [10.6](#) Hierarchical prediction structure
- [10.7](#) Quality Scalability
- [10.8](#) Spatial scalability, two layers
- [10.9](#) Medium Grain Quality Scalability
- [10.10](#) Combined Spatial, Temporal and Quality scalability
- [10.11](#) Simulcast vs. scalable bitrates
- [10.12](#) Quality scalability at CIF resolution
- [10.13](#) Spatial + temporal scalability, CIF → 4CIF resolution
- [10.14](#) Spatial scalability, 720p → 1080p resolution
- [10.15](#) Three views of the same scene
- [10.16](#) Multiview video: view examples
- [10.17](#) Multiview video: views and frames

- [10.18](#) Interview prediction of key frames
- [10.19](#) Inter-view prediction of all frames
- [10.20](#) Overview of configurable video codec
- [10.21](#) Reconfigurable Video Coding scenario
- [10.22](#) Fully Configurable Video Coding framework

# ***List of Tables***

- [2.1](#) Video frame formats
- [2.2](#) ITU-R BT.601-5 Parameters
- [2.3](#) HD display formats
- [3.1](#) SAE of residual frame after motion compensation, 16 × 16 block size
- [3.2](#) Probability of occurrence of motion vectors in sequence 1
- [3.3](#) Huffman codes for sequence 1 motion vectors
- [3.4](#) Probability of occurrence of motion vectors in sequence 2
- [3.5](#) Huffman codes for sequence 2 motion vectors
- [3.6](#) MPEG-4 Visual Transform Coefficient (TCOEF) VLCs: partial, all codes < 9 bits
- [3.7](#) MPEG4 Motion Vector Difference (MVD) VLCs
- [3.8](#) Motion vectors, sequence 1: probabilities and sub-ranges
- [3.9](#) Residual luminance samples: top-right 8 × 8 block
- [3.10](#) DCT coefficients
- [3.11](#) Quantized coefficients
- [3.12](#) Variable length coding example
- [3.13](#) Rescaled coefficients
- [3.14](#) Decoded residual luminance samples
- [4.1](#) Overview of the H.264 standard document
- [5.1](#) Syntax examples: format
- [5.2](#) H.264 Syntax Sections
- [5.3](#) Selecting a reference picture to re-map
- [5.4](#) Initial reference picture list
- [5.5](#) Final reference picture list
- [5.6](#) Selected NAL unit types
- [5.7](#) Sequence Parameter Set example
- [5.8](#) Picture Parameter Set example
- [5.9](#) Slice types in H.264
- [5.10](#) Slice Header, IDR/Intra, Frame 0
- [5.11](#) Slice Header, Inter, Frame 1
- [5.12](#) Macroblock types
- [5.13](#) coded block pattern examples.
- [5.14](#) Type and prediction elements for B macroblocks, excluding Direct or 8 × 8 partitions



- [5.15](#) I macroblock, example 1
- [5.16](#) P macroblock, example 2
- [5.17](#) P macroblock, example 3
- [5.18](#) P macroblock, example 4
- [5.19](#) B macroblock, example 5
- [5.20](#) B macroblock, example 6
- [5.21](#) B macroblock, example 7
- [5.22](#) B macroblock, example 8
- [6.1](#) Intra prediction types
- [6.2](#) Choice of prediction mode, most probable mode = 1
- [6.3](#) Reference picture sources
- [6.4](#) Reference frames and motion vectors for P and B macroblocks
- [6.5](#) Layers in hierarchical GOP example
- [7.1](#)  $V_i/4$  values,  $4 \times 4$  blocks
- [7.2](#) Table  $v$  defined in H.264 standard
- [7.3](#) Estimated Qstep ( $4 \times 4$  blocks) =  $\mathbf{V}_i/4 / \mathbf{S}_i \cdot 26$ , element-by-element division
- [7.4](#) Tables  $v$  and  $m$
- [7.5](#) 8-point Integer Transform Basis Cf8
- [7.6](#) 8-point Integer Transform Basis Cl8
- [7.7](#) Estimated Qstep ( $8 \times 8$  blocks)
- [7.8](#) Scaling matrices
- [7.9](#) Exp-Golomb Codewords
- [7.10](#) Mappings to code num
- [7.11](#) Choice of look-up table for coeff token
- [7.12](#) Thresholds for determining whether to increment *suffixLength*
- [7.13](#) Binarization of MVD magnitude
- [7.14](#) Context models for bin 1
- [7.15](#) Context models
- [8.1](#) Selected formats, frame rates and levels
- [8.2](#) HRD example 1: access unit sizes
- [8.3](#) HRD example 2: frame sizes
- [8.4](#) HRD example 3: frame sizes
- [8.5](#) Macroblock allocation map types
- [8.6](#) Switching from stream A to stream B using SP-slices
- [8.7](#) SEI messages
- [8.8](#) Video Usability Information: selected parameters

- [9.1](#) JM software directory structure
- [9.2](#) Selected parameter sections in the JM encoder configuration file
- [9.3](#) Section of trace file showing Sequence Parameter Set
- [9.4](#) 'Low Complexity' and 'Basic' configurations
- [9.5](#) 'Basic' plus options
- [9.6](#) 'Basic', 'Best Baseline', 'Medium'
- [9.7](#) Bitrate and delay constraints
- [10.1](#) Example: Quality Scalability
- [10.2](#) Example: Spatial + Temporal Scalability

# ***1***

## ***Introduction***

### **1.1 A change of scene**

#### **2000:**

Most viewers receive analogue television via terrestrial, cable or satellite transmission.

VHS video tapes are the principal medium for recording and playing TV programs, movies, etc.

Cell phones are cell phones, i.e. a mobile handset can only be used to make calls or send SMS messages.

Internet connections are slow, primarily over telephone modems for home users.

Web pages are web pages, with static text, graphics and photos and not much else.

Video calling requires dedicated videoconferencing terminals and expensive leased lines. Video calling over the internet is possible but slow, unreliable and difficult to set up.

Consumer video cameras, camcorders, use tape media, principally analogue tape. Home-made videos generally stay within the home.

#### **2010:**

Most viewers receive digital television via terrestrial, cable, satellite or internet, with benefits such as a greater choice of channels, electronic programme guides and high definition services. Analogue TV has been switched o