Thomas Wiegand
and Gary J. Sullivan

# The H.264/AVC Video Coding Standard

The H.264/MPEG-4 Advanced Video Coding standard (H.264/AVC) is the newest video coding standard jointly developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). H.264/AVC has achieved a significant improvement in compression performance compared to prior standards, and it provides a network-friendly representation of the video that addresses both conversational (video telephony) and nonconversational (storage, broadcast, or streaming) applications. This article provides a description of the structure, technology, performance, and resources of H.264/AVC, which is referred to formally as ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10).

## BACKGROUND

Since the early 1990s, when video coding technology was in its infancy, international standards such as H.261, MPEG-1, H.262/MPEG-2 Video, H.263, and MPEG-4 Part 2 have been powerful engines behind the commercial success of digital video. They have played a pivotal role in establishing the technology by ensuring interoperability among products developed by different manufacturers. At the same time, these standards have allowed flexibility for optimizing and molding the technology to fit various applications, and for making cost-performance tradeoffs for particular product requirements. They have provided much-needed assurance to the content creators that their content will play everywhere, making it unnecessary to create and manage multiple copies of the same content to match the products of different manufacturers. Moreover, these standards have permitted economies of scale to allow steep reductions in cost for mass-market affordability. The latest addition to the lineup of these well-known standards is H.264/AVC.

## *MOTIVATION*

As in the case of other international video compression standards, the driving force behind the creation of H.264/AVC was the need to enable *interoperability* between encoder and decoder products made by different manufacturers while minimizing the quantity of encoded data necessary to achieve a given level of output video quality—a concept known as the *coding efficiency* of the design. In particular, the increasing demand for video services and the growing popularity of higher-definition video are constantly creating greater demand for improved compression capabilities and this, in turn, motivated the H.264/AVC standardization effort. As a result of advances in technology since the development of the prior standards, a substantial improvement in coding efficiency had become possible.

## *OBJECTIVES*

The main objectives of the H.264/AVC standard are focused on coding efficiency, architecture, and functionalities. More specifically, an important objective was the achievement of a substantial increase (roughly a doubling) of coding efficiency over MPEG-2 Video for high-delay applications and over H.263 version 2 for low-delay applications, while keeping implementation costs within an acceptable range. Doubling coding efficiency corresponds to halving the bit rate necessary to represent video content with a given level of perceptual picture quality. It also corresponds to doubling the number of channels of video content of a given quality within a given limited bit-rate delivery system such as a broadcast network. The architecture-related objective was to give the design a "network-friendly" structure, including enhanced error/loss robustness capabilities, in particular, which could address applications requiring transmission over various networks under various delay and loss conditions. The functionalities-related objectives included—as with prior video coding standards—providing support for random access (i.e., the ability to start decoding at points other than the beginning of the entire stream of encoded data) and "trick mode" operation (i.e., fast-forward, fast and slow reverse play, scene and chapter skipping, switching between coded bitstreams, etc.), and other features.

## *ISSUING BODIES AND SCHEDULE*

H.264/AVC was developed by the ITU/ISO/IEC Joint Video Team (JVT), consisting of experts from ITU-T's VCEG and ISO/IEC's MPEG organizations. VCEG is officially referred to as ITU-T SG16 Q.6, and it is a part of the Telecommunication Standardization Sector of the International Telecommunications Union (ITU-T, which is a United Nations organization for telecom-related standardization). MPEG is officially referred to as ISO/IEC JTC 1/SC 29/WG 11, and it falls jointly under the International Organization for Standardization and the International Electrotechnical Commission (ISO and IEC, which are major privately organized international standardization bodies).

In early 1998, VCEG issued a call for proposals on a project then called H.26L. The first draft design for that new standard was adopted in August 1999. In December 2001, VCEG and MPEG formed the JVT with the charter to finalize the



1053-5888/07/$25.00©2007IEEE

draft new video coding standard, which was formally approved as H.264/AVC in May 2003. A first extension was issued in September 2004 with version 3. The second major extension is expected to be finalized in early 2007 with version 7.

### TARGET APPLICATIONS

The applications foreseen for the H.264/AVC standard include broadcast over cable, satellite, cable modem, x (of any type) digital subscriber line (xDSL), and terrestrial channels; interactive or serial storage on optical and magnetic devices such as DVDs; storage and distribution of professional film and video material for content contribution, content distribution, studio editing, and post processing; video-on-demand or multimedia streaming services over cable modem, xDSL, local area network (LAN), integrated service digital network (ISDN), and wireless networks; conversational services over Ethernet, LAN, xDSL, ISDN, wireless and mobile networks, and modems; and multimedia messaging services over xDSL, Ethernet, LAN, ISDN, wireless, and mobile networks. With such broad application coverage, H.264/AVC quickly received a great deal of recent attention from industry and found widespread standard system adoption as well as deployment in products.

### STRUCTURE OF THE STANDARD

As has been the case for all ITU-T and ISO/IEC video coding standards, only the *bit-stream format* (i.e., encoded data format) and the central *decoding process* have been standardized in the H.264/AVC specification. The standard defines the syntax, certain constraints on allowed combinations of syntax values, and the decoding process of the syntax elements to convert the encoded bitstream data into a timed series of picture arrays of brightness and color values. As a result, every decoder conforming to the standard will produce the same output (prior to any postprocessing and display-oriented processes) when given an encoded bitstream as input, provided that the bitstream fully conforms to all of the constraints specified in the standard. However, all actual encoding algorithms, image acquisition, pre- and postprocessing operations, error/loss concealment and recovery, and all aspects of decoded video display, have been deliberately kept outside the scope of the standard. This limitation of the scope of the standard permits maximal freedom to optimize product designs in ways that do not interfere with interoperability (balancing compression quality, implementation cost, time to market, etc.). However, it provides no guarantees of video encoding quality or decoded video display quality. It allows encoders to produce any bitstream that is in the correct format. In fact, such a standard does not even require encoders to produce bitstreams that decode into video bearing any resemblance to their video input, or even require encoders to accept video input data at all. It also does not specify any particular relationship between the output of the specified decoding process and the output of the subsequent display process; it does not even specify that a decoder needs to have the ability to display its output.

## TECHNOLOGY

### FUNCTIONALITIES

The H.264/AVC technology design supports the coding of video for a wide variety of applications. In addition to enabling efficient compression of digital video, it supports error/loss resilience, random-access operation, "trick-mode" operation (mentioned earlier), region-of-interest preferential coding, stereo-view indicators, film-grain analysis/synthesis processing, and a variety of additional capabilities. Further work is underway to add enhanced application capabilities for scalable and multiview/three-dimensional video coding.

### ARCHITECTURE

The H.264/AVC design consists of a network abstraction layer (NAL) and a video coding layer (VCL). The NAL, which was created to fulfill the network-friendly design objective, formats data and provides header information for conveyance by transport layers or storage media. All data are encapsulated in *NAL units*, each of which contains an integer number of bytes. The NAL unit structure provides a generic form for use in both packet-oriented and bitstream-based systems. The format of NAL units is identical in both environments, except that each NAL unit is preceded by a unique start code prefix for resynchronization in bitstream-oriented transport systems. The VCL is specified to efficiently represent the content of the video data and fulfill the design objective of enhanced coding efficiency. It is similar in spirit to designs found in other standards in the sense that it consists of a hybrid of block-based temporal and spatial prediction in conjunction with scalar-quantized block transform coding. A simplified block diagram of typical encoder processing elements for the VCL is provided in Figure 1. Decoding processes are conceptually a subset of these encoding processes, and are shown in the shaded region of the figure.

Although only the decoding process is actually specified in the standard, we focus on typical encoder technology to explain it, since the design is much easier to understand from that perspective. As shown in Figure 1, the picture is split into blocks. The first picture at the start of a sequence or a random access point (a point within a coded video sequence at which effective decoding can begin) is typically coded in "intra" (intrapicture) mode, which means that only information from the picture itself is being used (no prediction references to other preceding pictures in the bitstream). Each sample of a block in such a picture is predicted using spatially neighboring samples of previously coded blocks in the same picture. For the remaining pictures of a sequence or between random access points, typically "inter" (interpicture) coding is utilized. Interpicture coding employs prediction (motion compensation) from other previously decoded pictures. The encoding process for interpicture prediction (motion estimation) includes selecting and encoding motion data. Such data identifies the reference picture and spatial displacements that are applied to the samples of each block.

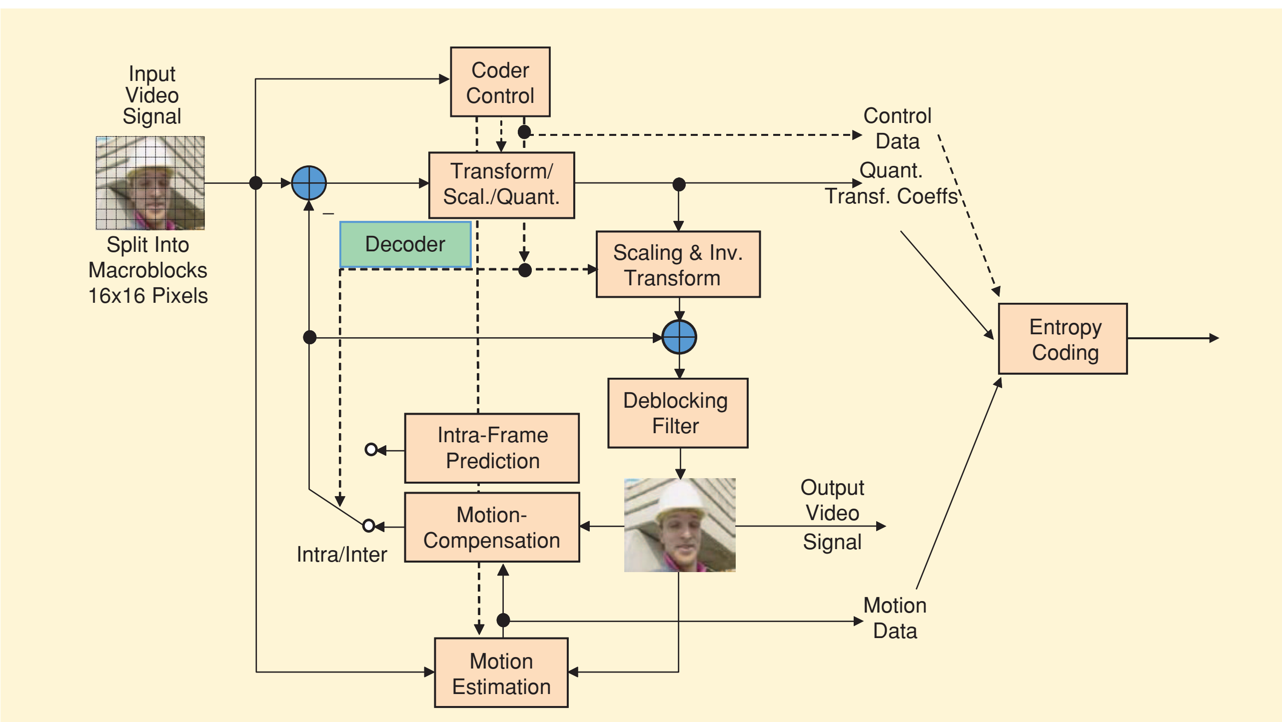
The residual difference computed as the error between each original and predicted (either an intrapicture or interpicture prediction) block is transformed using a decorrelating block transform. The transform coefficients are scaled and quantized. The quantized transform coefficients are entropy coded and transmitted together with the prediction selection information for either intra- or interpicture prediction.

The encoder duplicates much of the processing of the decoder to conduct prediction for the next blocks or pictures based on the same representation as will be decoded. Therefore, the quantized transform coefficients are inverse scaled and inverse transformed in the same way as in a decoder, resulting in the decoded prediction residual. This is then added to the prediction, and the result is processed by a deblocking filter, which finally yields the decoded video as its output.

### TOOLS

H.264/AVC contains a number of refinements and enhancements of prior coding tools, while making use of a similar basic structure. Some enhanced tools (not all of which are included in each profile of the standard, as mentioned later), ordered primarily in the order in which they might be likely to appear in the encoder's processing of an input picture, include:

- efficient interlaced video handling by either coding interlaced fields as distinct pictures or coding each 16 × 32 region using either field-based or frame-based coding techniques
- spatial-based intrapicture prediction with selectable block size and directional filtering
- enhanced motion-compensated interpicture prediction techniques, including quarter-sample motion, multiple reference picture use, variable block-size motion, weighted (scaled and offset) prediction, and highly flexible picture ordering and referencing ordering relationships
- 4 × 4 and 8 × 8 exact-match integer inverse transforms that are functionally similar to inverse discrete cosine transforms, with adaptive selection of the block size and a hierarchical second-level Hadamard transform for some dc components
- quantization step size control with logarithmically periodic step-size control and encoder-selectable frequency-dependent quantization scaling
- lossless modes for exact representation of some picture regions
- advanced entropy coding using context-adaptive binary arithmetic coding or context-adaptive variable-length coding
- a highly adaptive encoder-tunable deblocking filter applied before storing pictures as references for subsequent interpicture prediction (reducing visual artifacts and smoothing predictions)
- error/loss resilience features, including robust picture and sequence parameter set headers, flexible region ordering, arbitrary slice ordering, importance-separated data partitioning, redundantly coded picture regions, robust frame numbering, and robust picture order data
- special features for switching between similar bitstreams for such purposes as switching between streaming bit rates or "trick mode" playback operation
- a mechanism to ensure a lack of accidental emulation of specific




**[FIG1] Block diagram of typical encoding process for the VCL of H.264/AVC.**

"start code" patterns that are useful for systems-level multiplexing, error/loss recovery, and random-access functionality
- metadata carriage mechanisms to convey a variety of supplemental enhancement information
- auxiliary picture data for such purposes as alpha compositing or "object-oriented" video coding.

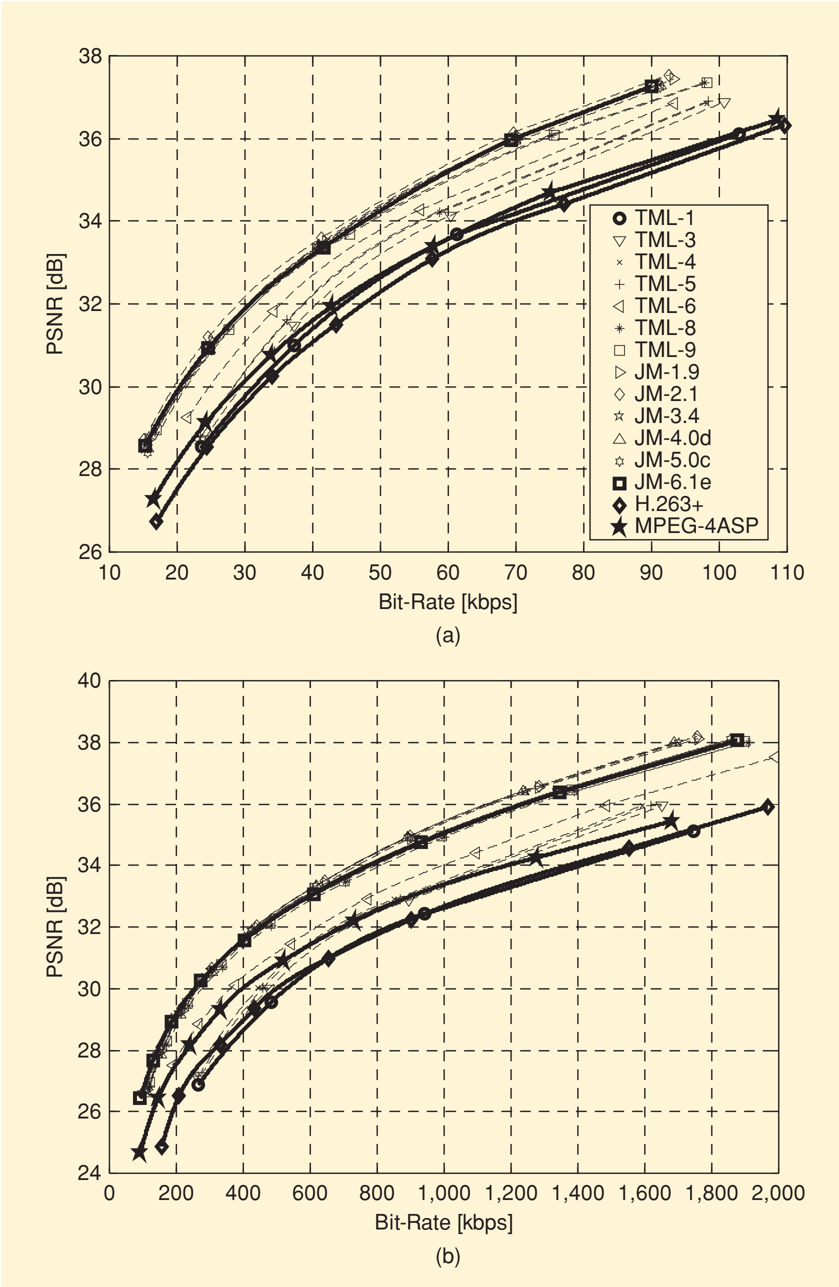
### PROFILES AND LEVELS

Profiles and levels specify well-defined sets of syntax constraints (for encoders) and decoder processing capabilities (for decoders), thereby enabling interoperability between encoder and decoder implementations within applications of the standard and between various applications that have similar functional requirements. A *profile* defines a set of syntax features for use in generating conforming bitstreams. A *level* places constraints on certain key parameters of the bitstream. All decoders conforming to a specific profile and level must support all features included in that profile when constrained as specified for the level. Encoders are not required to make effective use of any particular set of features supported in a profile and level, but they must not violate the syntax feature set and associated constraints. This implies, in particular, that conformance to any specific profile and level, although it ensures interoperability with decoders, does not provide any guarantees of video encoding quality.

Thus far, six profiles of H.264/AVC have been defined: Baseline, Extended, Main, High, High 10, and High 4:2:2. The Baseline profile is deployed in applications such as mobile phones, mobile TV, video telephony, and portable storage media such as the Apple video iPod. The Main profile is deployed in a limited number of storage media applications such as the Sony PlayStation Portable and SD TV. The High profile is deployed broadly for SD and HDTV (BSkyB, DirectTV, Dish Network, Premiere, etc.) and optical storage media (Blu-ray Disc and HD DVD). The Extended profile and High 10 profile have had no deployments reported yet. However, some recent interest has been expressed in High 10 by the studio/professional community, who are also the primary clients for the High 4:2:2 profile. Additionally, there is nearly completed work in the JVT for the professional community to add a new profile supporting the 4:4:4 color format and three new intra-only profiles supporting the 4:4:4, 4:2:2, and 4:2:0 color formats.

### COMPARISON WITH OTHER STANDARDS

H.264/AVC covers a broader range of applications than the well-known prior video coding standards, and it has substantially enhanced compression capability, substantially enhanced error/loss resilience capability, and greater flexibility for use in a broad




**[FIG2] Evolution of H.264/AVC from August 1999 to March 2003 exemplified for (a) QCIF sequence Foreman coded at 10 Hz and (b) CIF sequence Tempête coded at 30 Hz. The legend indicates the various versions of the test model, run with typical settings. All of the results shown were obtained using similar Lagrangian rate-distortion optimization methods.**

## H.264/AVC RESOURCES

**The Standard**

- ITU-T and ISO/IEC JTC 1, "Advanced Video Coding for Generic Audiovisual Services," ITU-T Rec. H.264 & ISO/IEC 14496-10, Version 1, May 2003; Version 2, Jan. 2004; Version 3 (with High family of profiles), Sept. 2004; Version 4, July 2005 [Online]. Available: http://www.itu.int/rec/T-REC-H.264

**Tutorials**

- G.J. Sullivan and T. Wiegand, "Video compression—From concepts to the H.264/AVC Standard," *Proc. IEEE*, vol. 93, no. 1, pp. 18–31, Jan. 2005.
- A. Luthra, G.J. Sullivan, and T. Wiegand, Eds., *IEEE Trans. Circuits Systems Video Technol. (Special Issue on the H.264/AVC Video Coding Standard)*, vol. 13, no. 7, July 2003.

**Overviews and Evaluations**

- G.J. Sullivan, "The H.264/MPEG4-AVC video coding standard and its deployment status," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP)*, Beijing, China, July 2005.
- T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan: "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Systems Video Technol.*, vol. 13, no. 7, pp. 688–703, July 2003.
- C. Fenimore, V. Baroncini, T. Oelbaum, and T.K. Tan, "Subjective testing methodology in MPEG video verification," in Proc. *SPIE Applications of Digital Image Processing*, Denver, CO, August 2004, vol. 5558, pp. 503–511.

**Books**

- I.E.G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*. New York: Wiley, 2003.
- S. Okubo, S. Kadono, Y. Kikuchi, and T. Suzuki: *H.264/AVC Textbook* (in Japanese, title translated). Tokyo: Impress Publisher, 2004.
- S. Ono, T. Murakami, and K. Asai, *Ubiquitous Technology: Hybrid Video Coding—MPEG-4 and H.264* (in Japanese, title translated). Tokyo: Ohmsha Press, 2005.

**Reference Software: (including various versions)**

- ITU-T and ISO/IEC JTC 1, "Reference software for advanced video coding," ITU-T Rec. H.264.2 & ISO/IEC 14496-5 (MPEG-4 Reference Software), 2005 (latest approved version) [Online]. Available: http://www.itu.int/rec/T-REC-H.264.2
- Draft versions are available for download from http://iphome.hhi.de/suehring/tml/

**Test Sequences and Bitstreams**

- ITU-T and ISO/IEC JTC 1, "Conformance specification for advanced video coding," ITU-T Rec. H.264.1 & ISO/IEC 14496-4 (MPEG-4 Conformance), 2005 (latest official version) [Online]. Available: http://www.itu.int/rec/T-REC-H.264.1
- MPEG industry forum resources [Online]. Available: http://www.mpegif.org/ resources.php

**Discussion List**

- Enrollment through http://mailman.rwth-aachen.de/mailman/listinfo/jvt-experts

**Resources for Further Development**

- JVT and VCEG committee documents [Online]. Available: http://ftp3.itu.int/av-arch
- H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable H.264/AVC extension," in *Proc. IEEE Int. Conf. Image Processing (ICIP'06)*, Atlanta, GA, Oct. 2006.
- T. Wiegand, G. J. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Joint draft 5: Scalable video coding," Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, Doc. JVT-R201, Bangkok, January 2006.
- K. Müller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, and T. Wiegand, "Multi-view video coding based on H.264/AVC using hierarchical B-frames," in *Proc. PCS 2006, Picture Coding Symp.*, Beijing, China, April 2006.

range of network environments. It is otherwise similar in purpose to prior well-known standards. Each of the features listed in the "Tools" section is substantially new in some key design aspect compared to most prior standards such as MPEG-2 Video. For example, most prior standards (with the exception of a profile of MPEG-4 Part 2) have not supported quarter-sample motion representation), no prior standard has included support for a block size smaller than $8 \times 8$ for motion representation blocks or residual transform blocks, and most prior video coding standards (with the exception of an optional mode of ITU-T Rec. H.263 version 2) have not specified the use of a deblocking filter.

## PERFORMANCE

### OBJECTIVE AND SUBJECTIVE QUALITY

As mentioned earlier, H.264/AVC has achieved a significant improvement in coding efficiency compared to prior standards. In Figure 2, we illustrate how the H.264/AVC capability improved through small steps over three to four years of design work for two example progressive-scan video sequences with typical coding options in publicly available reference software for drafts from August 1999 until completion in March 2003. The document and software versions have been called test model long-term (TML) during the earlier VCEG work and joint model (JM) when the work was continued in the JVT. As Figure 2 shows, the peak signal to noise ratio (PSNR) performance of TML-1 on these sequences was similar to that of H.263 version 2 and was below that of MPEG-4 Advanced Simple Profile (ASP). H.263 and MPEG-4 ASP were already significantly superior in compression capability to MPEG-2, so only those better-performing prior standards are shown in Figure 2. In comparison to TML-1, JM-6 achieved a relative performance improvement that is typically superior by 2-3 dB PSNR, or between 40 and 60% in bit rate reduction. A more detailed comparison of the coding performance of H.264/AVC relative to other standards and the Lagrangian optimization techniques used for comparing them can be found in the paper by Wiegand et al. cited in the "H.264/AVC Resources" sidebar. However, neither Figure 2 nor the paper mentioned above contain the improved techniques developed in the first extension (the High family of profiles) of H.264/AVC. The compression improvements shown using the objective PSNR measure have been confirmed using subjective quality assessment techniques.

### *SPEED/COMPLEXITY PERFORMANCE*

The computational resources necessary for encoding or decoding video using H.264/AVC and achieving its compression capability enhancement are substantially (perhaps roughly three times) higher than for prior video coding designs. However, the rapid adoption of the standard has shown that implementation costs have been kept within an acceptable range, considering the general easing of computational burdens provided by Moore's law since the prior standards were developed. Reducing the computational complexity of encoders and decoders for H.264/AVC has become an active area of research in the technical community.

**H.264/AVC Hardware and Software Products**

**Adopting Bodies and Consortia**

- 3GPP and 3GPP2 mobile environments
- ARIB (Japan), DMB (Korea), DAB (Europe), and DVB (Europe) broadcast standards
- AVC Alliance, IMTC, and MPEGIF promotional organizations
- Blu-Ray Disc Association and DVD Forum high-definition storage formats
- DLNA and ISMA multimedia streaming systems
- IETF audio-video transport real-time protocol payload packetization
- ITU-R broadcast and professional usage standards and ITU-T real-time multimedia conferencing systems
- MPEG storage file format and multiplex system standards
- NATO military specifications

**Hardware Products**

- Mobile phones by Nokia, Samsung, LG, etc.
- Broadcast HD encoders by Harmonic, Modulus, Scientific Atlanta/Cisco, Tandberg TV, Tut Systems, Thomson, etc.
- Direct-broadcast satellite deployments by BSkyB, DirecTV, Dish Network, Euro1080, Premiere, and ProSiebenSat.1.
- All HD DVD and Blu-Ray Disc players
- The Sony Playstation Portable and Apple iPod portable media player devices
- Videoconferencing systems by Polycom, Tandberg Telecom, etc.
- HD-capable decoding chips by Broadcom, Conexant, Micronas, Sigma Designs, and ST Microsystems

**Software Products**

- Media coder/player software packages, including Apple Quicktime, Cyberlink, Nero Digital, etc.
- DSP implementations for various processors
- SDK packages by Ateme, ATI, Elecard, FastVDO, Intel, KDDI, MainConcept, Sorenson, etc.
- Hardware accelerators for PC software decoders using graphics processing units (GPUs) by ATI, nVidia, etc.

## FURTHER TECHNICAL DEVELOPMENTS

A set of extensions for scalable video coding (SVC) is currently being designed in the JVT, aimed at the reconstruction of video signals with lower spatio-temporal resolution or lower quality from subsets of the coded video representation (i.e., from partial bitstreams.) Moreover, as an important side condition, SVC is additionally aimed at achieving a coding efficiency with the remaining partial bitstream that is comparable to that of "single-layer" H.264/AVC coding. In other words, the quality produced by decoding a subset of an SVC bitstream should be comparable to the quality produced by decoding an H.264/AVC bitstream that was encoded at the same bit rate in a nonscalable fashion. The SVC project will be finalized in early 2007. Another extension technology, that of efficient multiview video coding (MVC), has also been successfully demonstrated using H.264/AVC, requiring almost no change to the technical content of the standard. As a result, a standardized approach to MVC by extension of H.264/AVC is planned to be completed by the JVT in early 2008.

## RESOURCES

The ultimate resources on H.264/AVC are the standard text itself and its standardized conformance bitstreams and reference software. These are published by both ITU-T and ISO/IEC as so-called "twin texts" (i.e., they are technically aligned but published independently of each other). The most recent edition is from July 2005. Other resources for H.264/AVC are included in the "H.264/AVC Resources" sidebar.

## PRODUCTS

In addition to the video coding standard itself, a complete ecosystem has been created to enable its widespread use. Some application-domain and product uses of the new standard are outlined in the "H.264/AVC Hardware and Software Products" sidebar.

## AUTHORS

*Thomas Wiegand* (wiegand@hhi.de) is the head of the Image Communication research group of the Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute Berlin, Germany. He is the associate rapporteur of VCEG and a vice-chairman of the JVT and MPEG's video group. He has been the primary editor of the H.264/AVC standard.

*Gary J. Sullivan* (garysull@ microsoft.com) is a video architect at Microsoft Corporation of Redmond, Washington. He is the rapporteur of VCEG and a co-chairman of the JVT and MPEG's video group. [SP]