Extensions of ITU-T Recommendation H.324 for Error-Resilient Video Transmission

Niko Färber and Bernd Girod, University of Erlangen-Nuremberg
John Villasenor, University of California, Los Angeles

ABSTRACT The authors provide an overview of recent techniques that enable the use of ITU-T Recommendation H.324 with error-prone transmission channels. H.324 covers the overall architecture, algorithms, and syntax for low-bit-rate multimedia terminals. This overview focuses on extensions for error-resilient video transmission that are related to the H.263 video codec and H.223 multiplex. The latter is responsible for reliable delivery of data to each terminal component and is therefore of major importance to the overall system. In particular, the discussed approaches to robust video transmission require reliable performance of the multiplex scheme. The considered techniques for video coding are either compatible with the baseline mode of H.263 or part of new extensions recently devised by Study Group 16 as part of the H.263+ effort. This overview includes "Error Tracking" (Appendix II), "Reference Picture Selection" (Annex N), and "Independent Segment Decoding" (Annex R). Whenever appropriate, the description of those video extensions and multiplex schemes is complemented with simulation results for transmission over Rayleigh fading channels.

In the world of telecommunications, the technical superiority of a solution is not necessarily sufficient to guarantee its success. One reason for this observation is Metcalf's Law which states that the value of a network service grows with the square of the number of users. In other words, every new user brings additional value to everyone else connected to the network. For new services, such as mobile multimedia communication, Metcalf's Law results in the familiar "chicken-and-egg" start-up barrier, since the value of the first few terminals connected to the network is very low. One important factor that can encourage the investment necessary to surmount this barrier is the existence of a state-of-the-art international standard for the particular service. In general, the gain for an individual company contributing to such a standard outweighs the risk of sharing technical knowledge with potential competitors. Moreover, the joint effort of standardization often results in technically excellent solutions that combine interoperability with ample opportunity for innovative implementation and compatible improvements, and hence future product differentiation.

In the context of wireless video, it is therefore natural to look toward existing or emerging standards for video communication over wire lines. In this article we give a brief overview of International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) Recommendation H.324 which describes a complete multimedia terminal operating at low bit rates. While H.324 is intended for use over V.34 modems at a rate of 28.8 kb/s, it can be operated at rates as low as about 10 kb/s or (although without the V.34 modem) as high as several hundred kilobits per second. This makes it a strong candidate for emerging wireless multimedia systems, which will involve bit rates in the above range. However, certain extensions are necessary to cope with the increased error rate of wireless channels. The presentation of these extensions, with a focus on system components relevant to mobile video, is the main objective of this article.

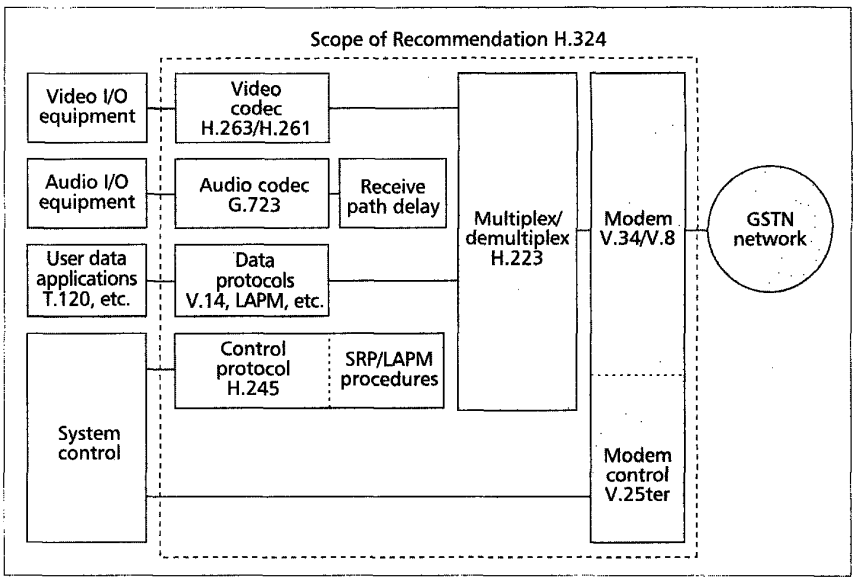
ITU-T H.324 Multimedia Terminal

ITU-T Recommendation H.324 [1] describes terminals for low-bit-rate multimedia communication, which may consist of real-time voice, data, and video, or any combination, including videotelephony. Because the transmission is based on V.34 modems operating over the widely available general switched telephone network (GSTN), H.324 terminals are likely to play a major role in future multimedia applications. In fact, several H.324-enabled products have already been developed and are being sold in rapidly increasing numbers.

As mobile communication becomes a more important part of daily life, the next step is to support mobile multimedia communication. Recognizing this development, the ITU-T started a new Ad Hoc Group (AHG) in 1994 to investigate the use of H.324 in mobile environments. This group, which is now formally part of ITU-T Study Group 16, Question 11, also initiated work on error-resilient video coding which is now continued in ITU-T Study Group 16, Question 15. In this article, we will refer to these respective groups as the Mobile and Video AHGs, and give an overview of the main issues discussed until now. The focus is on the video codec and multiplexing scheme, which are discussed in more detail in the third and fourth sections, respectively. The following overview of the H.324 system is based on [2] and [3], which are good introductions from the editor of the Recommendation. Whenever appropriate we point out relevant issues for mobile terminals.

Figure 1 illustrates the major components of an H.324 terminal. Only the modem (V.34 [4]), multiplex (H.223 [5]), and control protocol (H.245 [6]) are mandatory elements.

0163-6804/98/$10.00 © 1998 IEEE



■ **Figure 1.** *Components of an H.324 multimedia terminal.*

Although H.324 specifies the use of a V.34 modem, extension to other transmission systems is straightforward. The data, video, and audio streams are optional, which ensures that even the most basic H.324 terminals can interwork with more advanced terminals. During the setup of the communication, the terminals negotiate a common set of capabilities used for the session. This flexibility allows for a great range of possible products and applications, as well as future extensions. Applications include standalone videophones, PC-based multimedia applications, inexpensive voice/data modems, World Wide Web browsers with live video, video-only security cameras, and so on.

The original version of H.324 employs the V.34 modem directly as a synchronous data pump, to send and receive the bitstream generated by the H.223 multiplex. No data compression or retransmission protocols are used at the modem level, although these features can be invoked at a higher layer for some types of data streams. This view of the underlying transmission system simplifies substitution of the V.34 modem by some other connection-oriented transmission system without seriously affecting any other parts of the Recommendation. In particular, a "wireless interface" with a wide range of bit rates can be used instead of V.34 to transmit the multimedia streams. This, however, requires certain changes in the setup procedures (e.g., for the establishment of a connection). The changes necessary to replace the V.34 modem with a wireless interface are covered in Annex C of H.324.

The H.223 multiplex interleaves video, audio, data, and control streams into a single bitstream, allowing highly dynamic allocation of bandwidth to the different channels. It consists of a lower multiplex layer, which actually mixes the different media streams, and a set of adaptation layers, which perform logical framing, sequence numbering, error detection, and error correction by retransmission as appropriate to each media type. H.223 is a connection-oriented multiplexer, which is designed to mix an arbitrary number of channels on a circuit-switched network. For low bit rates, however, typically no more than one video, audio, and data channel are used in each direction. H.223 is byte-oriented and provides near-zero multiplex delay at very good efficiency. However, H.223 is not very robust against errors since it is designed to work with V.34 modems providing low error rates. Especially when the modem is operated below the maximum data rate of 28.8 kb/s, which is possible in steps of 2.4 kb/s, V.34 connections provide a bit error rate (BER) typically less than 10^{-6} and are characterized by completely error-free transmission for many seconds with short burst errors. The typical error patterns for many wireless environments are very different and generally more demanding. Therefore, H.223 is not suitable for wireless applications, as discussed in greater detail in the fourth section.

There are many predefined control messages that an H.324 terminal has to understand and treat correctly. Those messages, like the capability exchange messages mentioned above, are defined in the H.245 multimedia system control protocol. The control messages are sent over a reliable link layer with error recovery by retransmission using either the V.42 link access procedure for modems (LAPM) or Simplified Retransmission Protocol (SRP). The H.245 message structure can easily be extended to support new features. For example, most of the mobile extensions treated in this article also require changes in the H.245 protocol. These changes will be mentioned when appropriate.

To support user data applications, like electronic whiteboards and computer application sharing, several data protocols are supported that provide reliable transmission (V.14, LAPM, etc.). The G.723 audio codec can achieve near-toll-quality speech transmission at either 5.3 or 6.3 kb/s. User data applications as well as the audio codec will not be considered further in this article.

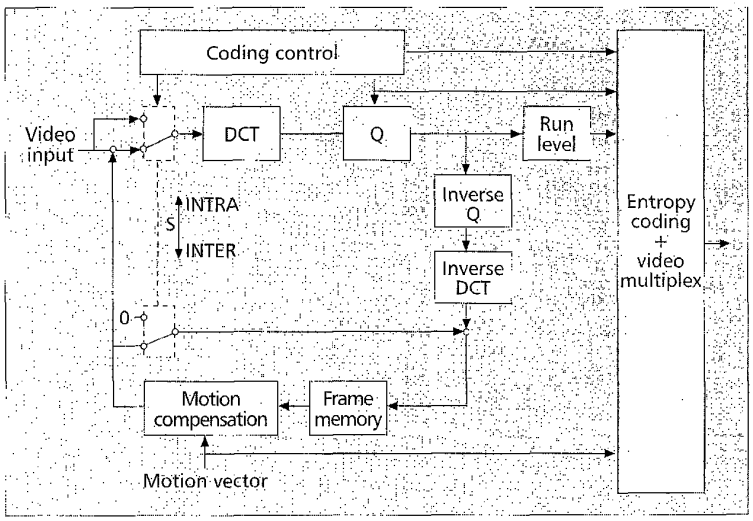
The remaining system component in Fig. 1 is the H.263 video codec. It was developed as an improvement over H.261, which is today's standard video codec for videoconferencing on the integrated services digital network (ISDN). H.263 can operate at a wide range of bit rates, but is particularly interesting at rates below 32 kb/s where acceptable quality is still possible for video containing limited motion. In the context of wireless multimedia, it is particularly important to consider the video codec, because video typically requires the largest portion of the total bit rate. Furthermore, motion-compensated prediction, which is one of the key components in video compression, causes spatiotemporal error propagation that is very annoying to end users. The mobile extensions discussed in the third section are therefore aimed at increased error robustness for the video codec.

ITU-T H.263 Video Codec

The Basic Concept of H.263

The basic concept of H.263 [7] is illustrated in Fig. 2. The video input is a sequence of digitized pictures (frames) at a rate of 30 frames/s with several possible picture resolutions. Although H.263 supports color video, we restrict our discussion to the luminance component. Furthermore, we focus on QCIF resolution (176 x 144 pixels), which is the most common input format for bit rates below 64 kb/s.

At QCIF resolution, each picture is divided into 11 x 9 macroblocks (MB, 16 x 16 pixels), which are further subdivided into four 8 x 8 blocks. Each MB is processed according to Fig. 2. Two basic modes of operation can be selected for each MB depending on the switch *S*. If the INTER mode is selected the MB is predicted from the previously reconstructed



■ **Figure 2.** *A simplified block diagram of an H.263 encoder.*

frame using motion compensation. Note that the output from the frame memory in Fig. 2 is identical to the decoded frames at the decoder for error-free transmission. Therefore, the same prediction can be formed at the encoder and decoder. The term *motion compensation* means that the current MB is predicted by a 16 x 16 block in the previous frame which is spatially shifted in accordance with a motion vector. In the ideal case, the remaining prediction error is negligible, and no more information need be transmitted. In general, however, the remaining prediction error is encoded using a discrete cosine transform (DCT) for each 8 x 8 block. The transform coefficients are quantized (Q) and encoded as a series of zero runs and quantizer levels. Finally, run-level pairs, motion vectors, and mode information are entropy-coded along with other side information, resulting in variable-length code words which are multiplexed to the video bitstream.

The second mode that can be used for each MB is the INTRA mode, in which no temporal dependency from previous frames is introduced and the picture is directly DCT coded. The choice between INTRA and INTER mode is not the subject of the recommendation and may be selected as part of the coding control strategy. During normal encoding, however, the INTER mode is preferred because of its superior coding efficiency. In other words, for the same number of bits transmitted, the INTER mode can usually provide significantly better picture quality than the INTRA mode.

Because the multiplexed bitstream consists of variable-length code words, a single bit error may cause a loss of synchronization and a series of invalid code words at the decoder. Even if resynchronization is achieved quickly, the subsequent code words are useless if the information is encoded differentially, as is the case for the motion vectors, for example. Therefore, H.263 supports optional group of blocks (GOB) headers. In QCIF format, a GOB consists of 11 MBs arranged in one row. When the GOB-header is present, fast resynchronization is guaranteed by a leading start code, and any dependencies from previous information in the current frame are avoided. Because all information within a correctly decoded GOB can be used independent of previous information in the same frame, the GOB is often used as the basic unit for decoding. Typically, either the complete information of a GOB is used or the entire GOB is discarded.

Lacking a better measure, video quality is usually represented by the peak signal-to-noise ratio (PSNR), defined as 10 $\log(255^2/\text{MSE})$, where MSE is the mean squared error between the original video input signal with a peak-to-peak range of 255 amplitude levels and the reconstructed video at the decoder. The PSNR is expressed in decibels and is usually calculated for each frame. Furthermore, the overall performance measure $\overline{\text{PSNR}}$ is calculated by time-averaging the PSNR values for each frame over a whole sequence. Although it is difficult to correlate the PSNR with subjective image quality, a difference of 1 dB generally corresponds to a noticeable difference for H.263 coded video. Furthermore, a loss of PSNR caused by transmission errors is usually more serious than the same loss caused by higher compression, since the resulting artifacts are concentrated in a part of the picture and hence more annoying.

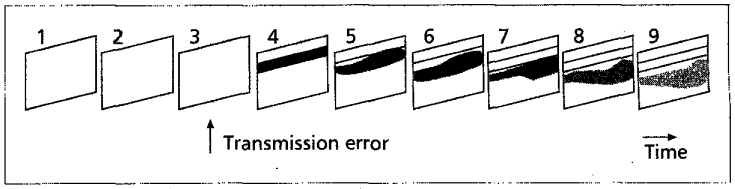
A number of minor and major enhancements in H.263 result in a significant performance gain over H.261. At 64 kb/s and 12.5 frames/s the gain in $\overline{\text{PSNR}}$ compared to H.261 can be close to 4 dB. Equivalently, for a given picture quality the bit rate can be reduced by a factor of two. One important enhancement of H.263 is the use of half-pel accuracy for motion compensation, which was limited to integer-pel accuracy in H.261. Furthermore, H.263 supports four optional encoding techniques ("options") which further improve performance at the cost of additional complexity. Those options are covered in Annexes D–G and have to be negotiated with the decoder via H.245. At 64 kb/s and 12.5 frames/s the gain in $\overline{\text{PSNR}}$ compared to the baseline mode of H.263 can be more than 2 dB when all options are employed. For a detailed performance analysis and description of H.263, see [8].

Transmission Error Effects

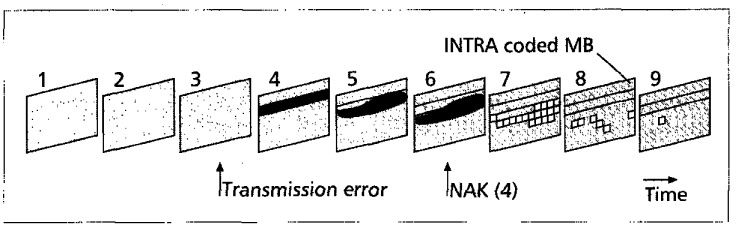
The compressed video signal is extremely vulnerable to transmission errors. In INTER mode, the loss of information in one frame has considerable impact on the quality of the following frames. As a result, temporal error propagation is a typical transmission error effect for predictive coding. Because errors remain visible for a longer period of time, the resulting artifacts are particularly annoying to end users. Figure 3 illustrates the typical transmission error effects for the loss of one GOB in frame 4. As indicated, the error may also propagate spatially from its original occurrence due to motion-compensated prediction. To some extent, the impairment caused by transmission errors decays over time due to leakage in the prediction loop. However, the leakage in standardized video decoders such as H.263 is not very strong, and quick recovery can only be achieved when image regions are encoded in INTRA mode (i.e., without reference to a previous frame). The INTRA mode, however, is not selected very frequently during normal encoding. In particular, no completely INTRA-coded frames are usually inserted in real-time encoded video as is done for storage or broadcast applications. Instead, only single MBs are encoded in INTRA mode for image regions that cannot be predicted efficiently.

The severity of errors can be reduced if concealment techniques are employed by the decoder to hide visible distortion. The aim of error concealment is to estimate the content of erroneously received image regions by exploiting the redundancy of the video signal. A summary of concealment techniques that have been proposed in the context of hybrid video coding can be found in [9]. However, even sophisticated concealment cannot totally avoid image degradation, and the accumulation of several small errors can also result in poor image quality. In the following, we assume that the image content of corrupted GOBs is replaced by simply repeating the

previous frame. This approach yields good results for sequences with little motion. However, severe distortions are introduced for image regions containing heavy motion. Note that H.263 does not define or require any particular error handling approach. Error handling in H.263, if used, is left to the H.263 implementors.



■ **Figure 3.** *Illustration of error propagation effects in H.263.*



■ **Figure 4.** *Illustration of error propagation effects when error tracking is applied.*

MOBILE EXTENSIONS TO H.263

Appreciating Metcalf's Law, the Mobile AHG acknowledged that the interworking of mobile H.324 terminals with terminals connected to the GSTN at a reasonable complexity and low delay is an essential requirement. This resulted in the decision to use the H.263 video codec and G.723 audio codec without alterations, because transcoding would be too complex. With this decision, many promising proposals were excluded from consideration. For example, the reordering of the H.263 bitstream into classes of different sensitivity (*data partitioning*) has been proposed to enable unequal error protection. Although this approach is known to be effective, the parsing and reassembling of the bitstream with added error protection cannot be implemented in a low-complexity low-delay transcoder. Two other interesting proposals, which do not share this drawback, could not be adopted by the ITU-T because they would have required minor modifications of the H.263 bitstream syntax. In January 1996, the syntax was already frozen, and neither the SUB-VIDEO approach [10] nor the NEWPRED approach [11] were included with the original H.263 at that time.

Given the above constraints, standard compatible extensions for robust video transmission are a valuable enhancement of H.324 for use in mobile environments. For this reason, the error-tracking approach described later was adopted by the ITU-T. Because of its compatibility, no technical changes were needed in H.263. However, an informative appendix (Appendix II) was added to explain the basic concept of the error-tracking approach. In addition, minor extensions of the H.245 control standard were necessary to include the additional control message videoNotDecodedMBs and to extend the capability structure.

Shortly after the original version 1 of H.263 was formally adopted by the ITU-T in March 1996 (i.e., reached *decision*), the ITU-T launched a new effort to further improve H.263 without changing the basic concept of block-based motion compensation. This short-term effort resulted in version 2 of the H.263 video coding standard, which is informally known as H.263+ and was adopted by the ITU-T in February 1998. It contains a number of optional feature enhancements which are added to the already existing options in an upward-compatible way in Annexes I–T. In particular, the two proposals for increased error robustness that could not be included in version 1 were slightly modified and enhanced during the development of version 2. As a result, the NEWPRED approach is included in Annex N (Reference Picture Selection), while the SUB-VIDEO approach is included in Annex R (Independent Segment Decoding) of the new Recommendation. These approaches are described in later subsections of this article. There has also been significant interest in variable-length codes (VLCs) which can be decoded in both forward and reverse directions. These reversible VLCs (RVLCs) offer the possibility to begin decoding in the forward direction and, on encountering an error, to proceed to the end of data and decode in the reverse direction [12]. RVLCs were adopted for coding of motion vector data as Annex D of H.263+. Future versions of H.263 may also include data partitioning that would allow the use of unequal error protection and improved resynchronization, already available in MPEG-4 [13].

Error Tracking — The error-tracking approach utilizes the INTRA mode to stop temporal error propagation but limits its use to severely impaired image regions only. During error-free transmission, the more effective INTER mode is utilized, and the system therefore adapts to varying channel conditions. Note that this approach requires that the encoder has knowledge of the location and extent of erroneous image regions at the decoder. This can be achieved by utilizing a feedback channel between transmitter and receiver.

The feedback channel is used to send negative acknowledgments (NACKs) back to the encoder. NACKs report the temporal and spatial location of GOBs that could not be decoded successfully and had to be concealed. For H.263, the temporal and spatial location of a GOB can be encoded by the time reference (TR, 8 bits) and group number (GN, 5 bits). The resulting rate on the feedback channel is then mainly determined by the GOB error rate, but will in general be very small.

Based on the information of a NACK, the encoder can reconstruct the resulting error distribution in the current frame (i.e., *track* the error from the original occurrence to the current frame). To do so, the encoder stores error-tracking information for the most recently encoded frames on a first-in first-out basis. Upon receipt of a NACK the encoder can use this information to reconstruct the error propagation process using an error-tracking algorithm. Note that the encoder itself "knows" all the information necessary for a perfect reconstruction of spatiotemporal error propagation. However, the storage and evaluation would be extremely demanding and infeasible for a practical system. Instead, a low-complexity algorithm can be used that provides a sufficiently accurate estimate of the true error distribution. Based on this estimate, the coding control of the encoder can be modified to stop error propagation effectively by selecting the INTRA mode whenever an MB is severely distorted. On the other hand, if error concealment was successful and the error of a certain MB is only small, the encoder may decide against INTRA coding. For more information on the error-tracking approach with a description of a low-complexity algorithm, the reader is referred to Appendix II of the H.263 video compression standard or [14]. However, it should be noted that the implementation of the error-tracking algorithm need not be standardized. Instead, evaluation of the NACKs is up to the

implementor, allowing considerable flexibility for the trade-off of accuracy vs. complexity.

Figure 4 illustrates the error-tracking approach for the same example as in Fig. 3. As soon as the NACK is received after a system-dependent delay, the impaired MBs are determined and error propagation can be terminated by INTRA coding these MBs (frames 7–9). In addition, Fig. 5 shows averaged simulation results for an assumed round-trip delay of 800 ms. The loss of picture quality (Δ PSNR) compared to the error-free case is calculated for each frame in the sequence. Compared to the case without error compensation, the picture quality recovers rapidly as soon as INTRA coded MBs are inserted into the bitstream. Note that the round-trip delay in this simulation is much greater than the maximum of 250 ms desirable for conversational services, which is all too common for very low-bit-rate transmission. It is also important to note that the error-tracking approach does not introduce any additional delay. A longer delay just results in a later start of the error recovery. Considering the slow recovery for "concealment only," NACKs may still be useful after several seconds.

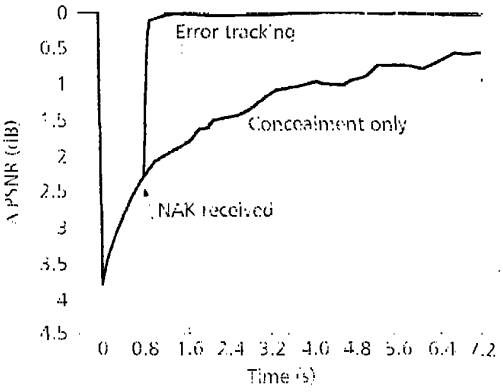


Figure 5. *Error recovery with feedback channel.*

Independent Segment Decoding — The independent segment decoding (ISD) mode is described in Annex R of H.263 and is therefore part of H.263+. In the following we assume that each GOB starts with a GOB header and that the "slice structure mode" (Annex K) is not in use, in which case a segment is identical to a GOB. In the ISD mode, each GOB is encoded as an individual picture (or subvideo) independent of other GOBs. In particular, all GOB boundaries are treated just like picture boundaries. This approach significantly reduces the efficiency of motion compensation, particularly for vertical motion, since image content outside the current GOB must not be used for prediction. We observed typical losses in PSNR in the range from 0.2 to 1.0 dB.

In the presence of transmission errors, the ISD mode ensures that errors inside a GOB will not propagate to other GOBs, as illustrated in Fig. 6. Of course, the ISD mode alone does not solve the problem of temporal error propagation. It only simplifies keeping track of the error effects. The error propagation itself must be combatted by feedback-based INTRA updates or, more effectively, by the use of the reference picture selection mode.

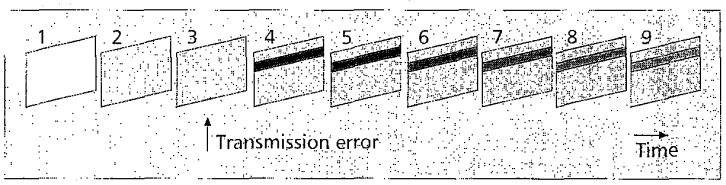


Figure 6. *Illustration of error propagation effects in the independent segment decoding mode.*

Reference Picture Selection — Similar to the error-tracking approach, the reference picture selection (RPS) mode of H.263+ also relies on a feedback channel to efficiently stop error propagation after transmission errors. This mode is described in Annex N, and is based on the NEWPRED approach mentioned above. Instead of using INTRA coding of MBs, the RPS mode allows the encoder to select one of several previously decoded frames as a reference picture for prediction. As for the discussion of the ISD mode, we again consider the case that all GOB headers are present. Then the reference picture is selected on a GOB basis; that is, for all MBs within one GOB the same reference picture is used. In order to stop error propagation while maintaining the best coding efficiency, the last frame available without errors at the decoder should be selected. The RPS mode can be combined with the ISD mode to avoid spatial error propagation or, for better coding efficiency, with an error-tracking algorithm.

RPS can be operated in two different modes. In the *ACK mode* all correctly received GOBs are acknowledged and the encoder only uses acknowledged GOBs as a reference. If the round-trip delay is greater than the encoded picture interval, the encoder has to use reference GOBs that are temporally far apart. This results in decreased coding performance for error-free transmission. In the case of transmission errors, however, only little fluctuations in picture quality occur. The second mode is called *NACK mode*. In this mode only erroneously received GOBs are signaled by sending NACKs. During error-free transmission, the operation of the encoder is not altered and the previously decoded GOBs are used for reference. After a transmission error, the decoder sends a NACK for erroneous GOBs

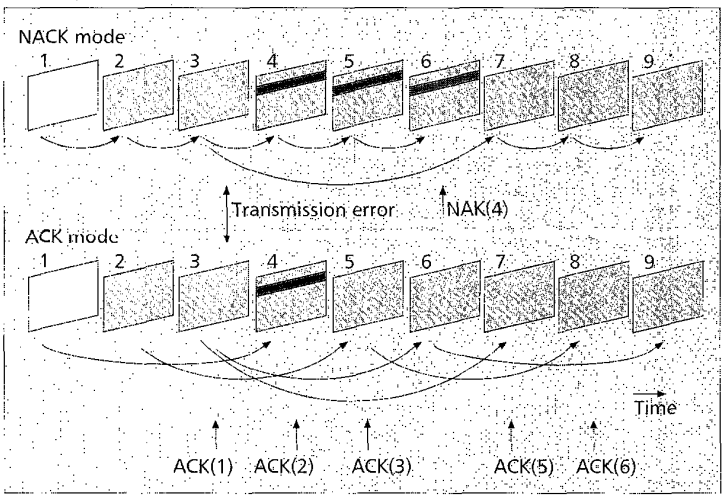


Figure 7. *Illustration of error propagation effects when the RPS mode is used in combination with the independent segment decoding mode. The arrows indicate the selected reference pictures for the corrupted GOB.*

and thereby requests that older frames be used as the reference GOB. In this mode longer quality degradations occur, since errors are permitted to propagate for the period of one round-trip delay. However, no performance loss during error-free transmission needs to be accepted. Therefore, the ACK mode is preferred in highly error-prone environments, while the NACK mode is advantageous if errors occur only rarely after longer periods of error-free transmission.

E_b/N_0 (dB)	BER	PER
20	0.002578	0.017075
22	0.001646	0.011250
24	0.001025	0.007575
26	0.000644	0.004675
28	0.000390	0.002775
30	0.000234	0.001725

■ **Table 1.** *Summary of test channel parameters.*

The typical transmission error effects for the RPS mode in both signaling modes are illustrated in Fig. 7, where the selection of reference GOBs is indicated by arrows. Note that the use of the ISD mode is assumed and the indicated selection is only valid for the erroneous GOB. In the NACK mode, the encoder receives a NACK for frame 4 before the encoding of frame 7. The NACK includes the explicit request to use frame 3 for prediction, which is observed by the encoder. In the ACK mode, no ACK is received for frame 4, and it is therefore never used for prediction.

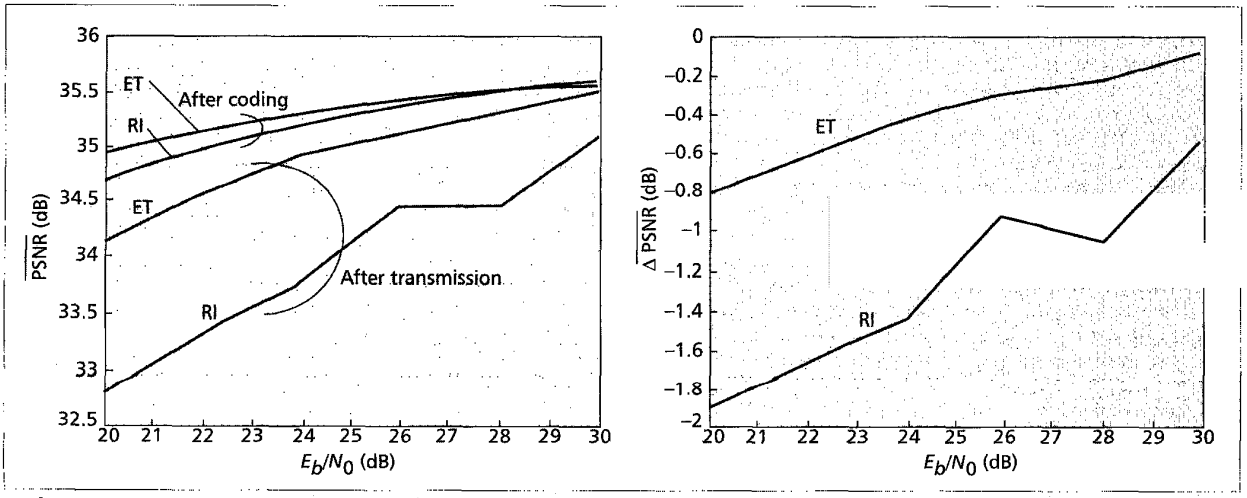
Note that the RPS mode requires additional frame buffers at the encoder and decoder to store several previous frames. Without the RPS mode only one frame has to be stored. This increased storage requirement may be a problem for inexpensive mobile terminals, but is certainly acceptable for PC-based systems when small picture resolutions are used. The advantage of the RPS mode as opposed to simply switching to INTRA mode lies in increased coding efficiency. Usually, fewer bits are needed for the motion-compensated prediction error than for the video signal itself, even if the time lag between the reference frame and the current frame is several frame intervals.

H.263 Transmission over Mobile Channels

In this section we investigate the performance of the discussed H.263 extensions in a mobile environment by simulating the transmission over a wireless digital European cordless telephony (DECT) channel. Our simulations are based on bit error sequences that are generated assuming Rayleigh fading and a velocity of 14 km/hr. Carrier-to-noise ratios (E_b/N_0) in the range of 20–30 dB were investigated, with corresponding BERs as summarized in Table 1. The bit error sequences exhibit severe burst errors and provide a total bit rate of 80 kb/s, which is the available bit rate in the double slot format of DECT.

For simplicity we ignore the H.223 multiplex for now, and apply forward error correction (FEC) directly to the video bitstream. We employ a BCH code of block size $n = 255$ bits with $k = 179$ information bits and $t = 10$ correctable errors/block. The measured packet error rates (PERs) for this FEC are also summarized in Table 1. The same code is used in all simulations. We compare the error-tracking approach which relies on NAKs transmitted over a feedback channel to simple "random-INTRA" updating. For error tracking, the feedback channel is assumed to be error-free with a constant delay of 100 ms. The round-trip delay, measured from encoding a frame to receiving a NAK, is about 300 ms. With the random-INTRA approach, N_I MBs were encoded in INTRA mode in each frame, where N_I was selected to match the average number of INTRA-coded MBs in the case of error tracking. The selection of individual MBs, however, is not based on feedback information but done randomly.

Twelve seconds of a typical videophone sequence (*mother and daughter*) are encoded at 12.5 frames/s and transmitted over 30 different realizations of each test channel. Figure 8 shows the picture quality measured by $\overline{\text{PSNR}}$ after coding (error-free) and after transmission on the left side. In addition, the loss of picture quality $\overline{\Delta\text{PSNR}}$ is provided on the right. The results are averaged values for all frames and realizations at each E_b/N_0. Note that the picture quality after coding increases with increasing E_b/N_0 because fewer NAKs are received at better channel conditions, and fewer MBs have to be coded in INTRA mode. The error-tracking approach performs significantly better than the random-INTRA update scheme. Alhough no simulation results are provided for the RPS mode, we remark that its performance in terms of $\overline{\Delta\text{PSNR}}$ is similar to



■ **Figure 8.** *Picture quality after coding and transmission (left) and loss of picture quality (right). The error-tracking and random-INTRA approaches are denoted ET and RI, respectively. The test sequence is mother and daughter.*

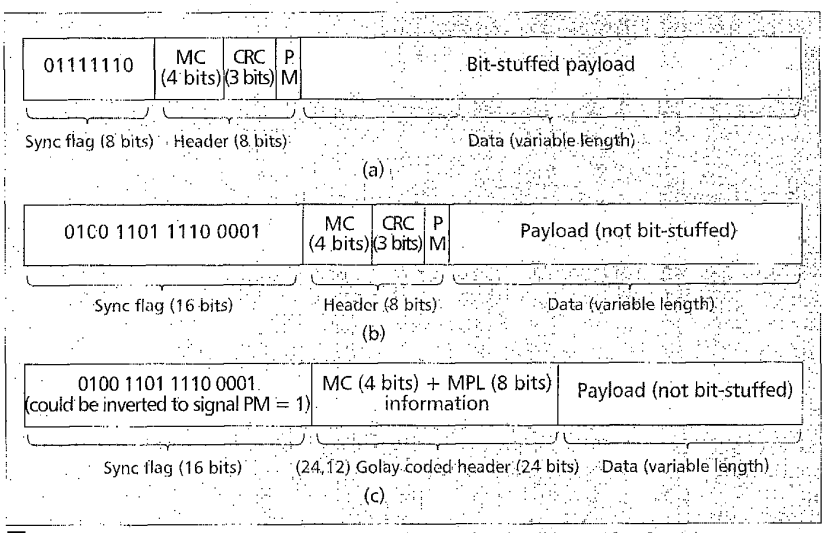


Figure 9. *H.223 packet structure for level 0 (a), level 1 (b), and level 2 (c).*

error-tracking. However, the picture quality after coding $\overline{\text{PSNR}}$ is somewhat better for lower E_b/N_0 since the INTRA mode can often be avoided. For very good channel conditions, however, the $\overline{\text{PSNR}}$ for error-tracking is superior.

ITU-T H.223 Multiplex

The effects of channel errors on the quality of a wireless video session are a function not only of the steps taken in the video coder and decoder but also of the other layers of the communication system. In the previous section, we applied bit error patterns from the channel directly to the video bitstream, which is in turn sent to the input of a video decoder. This simplified approach is very common, and can indeed be useful for exploring the value of different codec design approaches. However, it does not account for the important effects of the other communication layers on the video. While it is impossible at present to identify or fully characterize all the different wireless networks and systems which will eventually carry video traffic, it is certain that these networks themselves will provide a very diverse spectrum of error handling methods. For those networks that use protocols to ensure that data (in this case, video) is delivered error-free, application of channel errors directly to the compressed video in simulations is unreasonably pessimistic, even though late packets may pose a problem. For those networks that do not use retransmission, application of raw channel errors to encoded video fails to account for any FEC performed at other network layers. Multiplexing and/or packetization constitutes another important source of error because of the possibility that video can be misdelivered, causing large chunks of data to disappear from the video bitstream seen by the receiver. Similarly, an error at the packet level can cause large amounts of nonvideo to be directed to a video decoder.

The role of other layers of the communication system in video quality raises many challenges. On one hand, it is desirable to maintain a clean separation of layers, and to be able to design and optimize the video codec independent of other aspects of the communication systems. On the other hand, the nature of the errors that will be seen at the video level are very highly dependent on other aspects of the communication system and link. With the exception of systems that guarantee near-perfect delivery of video, therefore, some consideration of layers external to video seems inevitable in video codec design.

Probably the most extensive effort to date to jointly consider wireless video communications in the context of an end-to-end system has been performed in connection with recent work of the Mobile AHG to develop error-resilient enhancements to H.324 and H.223 in particular. While the original H.223 specification targeted the V.34 modem and was therefore designed with relatively low error rates in mind, interest in using H.324 over wireless channels led to work to extend H.223 to allow operation over error-prone channels. This work, which was carried out in large part during the period 1995–1997, led to the development of a series of annexes to H.223. With the addition of these annexes, Recommendation H.223 now offers a hierarchical, multilevel multiplexing structure allowing implementers to trade off robustness against overhead and complexity. In the simplest default layer of H.223 (level 0) which corresponds to the original specification developed for the V.34 modem, packets are variable-length and delimited by an 8-bit synchronization flag. A synchronization flag is followed by an 8-bit header containing a multiplex code (MC), which identifies the contents of the packet, and then by the payload, which in general can contain a mix of various sources. The end of the packet is indicated by the next appearance of the 8-bit synchronization flag. Bit stuffing is performed on all data between synchronization flags to avoid flag emulation. Figure 9a illustrates the H.223 level 0 packet structure. The principal vulnerabilities of H.223 level 0 lie in the bit stuffing, and in the short and therefore vulnerable synchronization flags and headers.

In level 1 (Fig. 9b) bit stuffing is not performed, and a longer synchronization flag is used. The flag can be emulated by the data, but such emulations are not usually problematic. In level 2 (Fig. 9c) further robustness is enabled by lengthening and adding error protection to the header that describes the contents of the packets. There is also a higher level (level 3), which involves FEC performed by the multiplexer on the data contained within the packets. The contents of the payload in Fig. 9 are identified by the MC. The packet marker (PM) is a 1-bit field that concerns the relationship between payloads in successive packets. In level 2 a multiplex payload length (MPL) field is provided that identifies the length of the payload in bytes. This provides additional redundancy by informing the receiver where the next sync flag can be expected. For levels 1 and 2, the multiplex packet length is an integer number of bytes. By contrast, in level 0 the bit stuffing will generally result in packets that are not byte-aligned.

Table 2 illustrates the performance of Levels 0 through 2. The table considers the ability of the H.223 multiplexer levels to deliver packets over a Rayleigh channel. The same basic channel model as in the previous section is used, however, with different channel parameters. A lower velocity of 1.4 km/hr is assumed, and the investigated Ratios E_b/N_0s are 14, 18, and 22 dB, respectively. In the simulations used to construct these tables, it was assumed that the channel errors derived from the channel model were applied directly to the multiplexed bitstream. The size of the packets was randomly chosen according to a uniform distribution between 1 and 125 bytes of payload data, excluding the overhead due to synchronization and data headers. A total of X = 100,000 packets were used in each simulation. Y_1 represents the number of packets that arrive at the receiver with

E_b/N_0 (dB)	BER	Level	Y_1/X	Y_2/X	Y_1+Y_2	Throughput
14	9.3×10^{-3}	0	90.33%	3.57%	93894	89.57%
		1	92.02%	2.65%	94672	91.30%
		2	92.93%	0.07%	93112	93.14%
18	3.7×10^{-3}	0	94.97%	2.51%	97480	94.37%
		1	95.83%	2.02%	97856	95.27%
		2	96.03%	0.07%	96217	96.07%
22	1.5×10^{-3}	0	97.72%	1.30%	99025	97.39%
		1	98.74%	0.47%	99212	98.64%
		2	97.42%	0.07%	97625	97.45%

■ **Table 2.** *Performance of levels 0–2 of ITU-T H.223 for Rayleigh fading channels. X: total number of packets transmitted (100,000 in the simulations); Y_1: number of packets received with error-free headers; Y_2: number of packets received in which the header contains undetected errors. Throughput measured in units of bits vs. packets; expresses the percentage of information bits delivered (perhaps with channel errors) to the correct decoder (video, audio, etc.).*

no errors in the header field. The payload contents of the Y_1 packets will therefore be delivered to the correct decoders (voice, video, etc.) for further processing, although the payload bits themselves may in general have channel-induced errors. Y_2 represents the number of packets that arrive with undetected header errors. This is a particularly damaging type of error, because it will cause the demultiplexer logic to incorrectly conclude that the multiplex header is uncorrupted, and then to distribute the payload data incorrectly to the source decoders. The table gives the ratios Y_1/X and Y_2/X expressed as percentages as well as the sum $Y_1 + Y_2$ expressed in absolute terms. Note that quantity $X - Y_1 - Y_2$, which is not shown in the table but can be trivially calculated, represents the number of packets in which header errors occur and are detected, thereby giving the demuliplexer the opportunity to protect the video and audio decoders from being fed garbage input. The final statistic in the table is the throughput, expressed as the ratio of the total number of received payload bits in the Y_1 packets with error-free headers to the total number of transmitted payload bits. If all packets were of equal length, the throughput would be equal to Y_1/X. The throughput is expressed in terms of bits instead of packets, and is useful in answering questions such as "what fraction of the transmitted video bits will be delivered (although perhaps with channel errors) to the video decoder?" As expected, the table shows that more robust multiplexer levels lead to improved communication. The degree of improvement is greater for poorer channels (i.e, lower E_b/N_0). Among the categories considered, the most important improvement as the multiplexer level is increased is in the percentage of data delivered to the decoder that is corrupted (Y_2/X in the table). Significantly, the most robust level of the multiplexer (level 2) reduces the amount of corrupted data by over an order of magnitude.

In more general terms, Table 2 shows that channel-error-induced failures at other network layers are likely to have extremely important consequences at the layers where the source codecs are located. For example, it is quite unlikely that an H.324 system designed for wireline environments (and therefore using H.223 level 0) would function over the given channels. Even a video codec modified to be extremely robust would be of only marginal use in a system in which a few percent of the video bits are misdelivered (e.g., to an audio decoder), creating large gaps in the received bitstream seen by the video decoder. As wireless video communications become a commercial reality over the coming years, it will be extremely important for designers to adopt a system-level view of video communications, and to avoid, for example, assuming that a third-party-supplied video codec designed for a reliable channel will function well for wireless applications.

Conclusions

The ITU-T H.324 multimedia terminal standard is likely to play a major role in the introduction of mobile multimedia communication, including mobile video. The error-prone channels typical for mobile communications require additional extensions to the H.324 system to achieve acceptable performance. This has been the task of the Mobile and Video AHGs of ITU-T Study Group 16 since 1994. An overview of the current status of these extensions with a focus on the H.263 video codec and the H.223 multiplex has been presented in this article. The extensions to the H.223 multiplex allow a trade-off between error robustness and complexity, and in some conditions can reduce the amount of corrupted data by over an order of magnitude. The extensions to the H.263 video codec are either compatible with the baseline mode of H.263 or included as options in H.263+. In both cases, the usage of feedback information can improve robustness significantly. Compatibility with the baseline mode, as provided by the error-tracking approach, is particularly important for interworking with existing H.324 terminals that cannot support H.263+. Work on error-robust video is continuing in both the ITU and MPEG-4, and we expect that error-handling techniques will become major considerations in the design of future multimedia terminals.

Acknowledgments

We gratefully acknowledge the work of all participants in the ITU efforts to investigate and standardize the error-resilient extensions for H.324 described in this article. Any "transmission errors" that might have occurred through simplification or emphasis of certain aspects over others are, of course, the sole responsibility of the authors.

References

[1] ITU-T Rec. H.324, "Terminal for Low Bitrate Multimedia Communication," 1995.

[2] D. Lindberg, H. Malvar, "Multimedia Teleconferencing with H.324," *Standards and Common Interfaces for Video Information Systems*, K. R. Rao, Ed., Bellingham, WA: SPIE Optical Engineering Press, 1995, pp. 206–32.

[3] D. Lindberg, "The H.324 Multimedia Communication Standard," *IEEE Commun. Mag.*, vol. 34, no. 12, pp. 46–51, Dec. 1996.

[4] ITU-T Rec. V.34, "A Modem Operating at Data Signaling Rates of Up to 28,800 bit/s for Use on the General Switched Telephone Network and on Leased Point-to-Point 2-Wire Telephone-Type Circuits," 1994.

[5] ITU-T Rec. H.223, "Multiplexing Protocol for Low Bitrate Multimedia Communication," 1996.

[6] ITU-T Rec. H.245, "Control Protocol for Multimedia Communication," 1996.

[7] ITU-T Rec. H.263, "Video Coding for Low Bitrate Communication," 1996.

[8] B. Girod, N. Färber, and E. Steinbach, "Performance of the H.263 Video Compression Standard," *J. VLSI Signal Processing: Sys. for Signal, Image, and Video Tech.*, vol. 17, Nov. 1997, pp. 101–11.

[9] J.W. Park, J.W. Kim, and S.U. Lee, "DCT Coefficient Recovery-Based Error Concealment Technique and its Application to MPEG-2 Bit Stream Error," *IEEE Trans. Circuits and Sys. for Video Technology*, vol. 7, no. 6, Dec. 1997, pp. 845–54.
[10] LBC Doc. LBC-95-309 (ITU-T SG 15, WP 15/1), "Sub-videos with Retransmission and Intra-Refreshing in Mobile/Wireless Environments," National Semiconductor Corp., Darmstadt, Germany, 1995.
[11] LBC Doc. LBC-96-033 (ITU-T SG 15, WP 15/1), "An Error Resilience Method Based on Back Channel Signalling and FEC," Telenor Research, San Jose, CA, 1996.
[12] G. Wen and J. Villasenor, "A Class of Reversible Variable Length Codes for Robust Image and Video Eoding," *IEEE Int'l. Conf. Image Proc.*, vol. 2, 1997, pp. 65–68.
[13] R. Talluri, "Error Resiliency in ISO's MPEG-4 Video Coding Standard," *IEEE Commun. Mag.*, this issue.
[14] E. Steinbach, N.Färber, and B. Girod, "Standard Compatible Extension of H.263 for Robust Video Transmission in Mobile Environments," *IEEE Trans. Circuits and Sys. for Video Tech.*, vol. 7, no. 6, Dec. 1997, pp. 872–81.

BIOGRAPHIES

NIKO FÄRBER (faerber@nt.e-technik.uni-erlangen.de) received a Diplom-Ingenieur degree in electrical engineering from the Technical University of Munich, Germany, in 1993. He was then with the research laboratory Mannesmann Pilotentwicklung, where he developed system components for satellite-based vehicular navigation. In 1994 he joined the Telecommunications Laboratory at the University of Erlangen-Nuremberg and is now a member of the Image Communication Group. He started his research on robust video transmission as a visiting scientist at the Image Processing Laboratory of the University of California, Los Angeles. Since then he has published several conference and journal papers on the subject and has contributed successfully to the ITU-T Study Group 16 efforts for robust extensions of the H.263 standard. He also served as publicity vice chair for ICASSP '97 in Munich. Further research interests are combined source and channel coding and software-only video coding.

BERND GIROD [F] (girod@nt.e-technik.uni-erlangen.de) is a chaired professor of telecommunications in the Electrical Engineering Department of the University of Erlangen-Nuremberg, Germany. He received his M. S. degree in electrical engineering from Georgia Institute of Technology in 1980, and his doctoral degree with highest honors from the University of Hannover, Germany, in 1987. Until 1987 he was a member of the research staff at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, working on moving image coding, human visual perception, and information theory. In 1988 he joined Massachusetts Institute of Technology, Cambridge, first as a visiting scientist with the Research Laboratory of Electronics, then as an assistant professor of media technology at the Media Laboratory. From 1990 to 1993 he was professor of computer graphics and technical director of the Academy of Media Arts in Cologne, Germany, jointly appointed with the Computer Science Section of Cologne University. He was a visiting adjunct professor with the Digital Signal Processing Group at Georgia Institute of Technology, Atlanta, in 1993. Since 1993 he has been a professor of electrical engineering/telecommunications at the University of Erlangen-Nuremberg, Germany, and head of the Telecommunications Institute I. He served as chair of the Electrical Engineering Department from 1995 to 1997 and has served as director of the Center of Excellence "3-D Image Analysis and Synthesis" since 1995. His research interests include multidimensional signal processing, information theory, video signal compression, human and machine vision, sensory computing, computer graphics and animation, as well as interactive media.

JOHN VILLASENOR (villa@icsl.ucla.edu) received a B.S. degree from the University of Virginia in 1985, an M.S. from Stanford University in 1986, and a Ph.D. from Stanford in 1989, all in electrical engineering. From 1990 to 1992 he was with the Radar Science and Engineering section of the Jet Propulsion Laboratory in Pasadena, California, where he developed interferometric terrain mapping and classification techniques using synthetic aperture radar data. He joined the University of California, Los Angeles in 1992 and is currently associate professor and vice chair of the Electrical Engineering Department. His current research interests are in joint source and channel coding, wireless multimedia communications, concatenated codes, and low-complexity image and video coding architectures and algorithms. He has also created a research program in configurable computing architectures, which involves machines that modify their hardware over millisecond time scales in order to track changing characteristics of the data or processing environment. He serves as an associate editor of *IEEE Transactions on Signal Processing* and *IEEE Transactions on Circuits and Systems for Video Technology*. He has also served as co-chair for the ad hoc subgroup on wireless multimedia within the ITU activity in very-low-bit-rate visual telephony, He was a proposer, in collaboration with Texas Instruments, of the set of MPEG 4 core experiments to evaluate robust variable-length codes, and of the robust variable-length codes adopted for H.263+. He is a program committee member of the IEEE Data Compression Conference and a member of the IMDSP Technical Committee within the IEEE Signal Processing Society.