Quadtree-Based Reconfigurable Cordless Videophone Systems

Jürgen Streit and Lajos Hanzo, *Senior Member, IEEE*

Abstract—**Arbitrarily programmable, but fixed-rate quadtree (QT) decomposed, parametrically enhanced videophone codecs using quarter common intermediate format (QCIF) video sequences are proposed as a direct replacement for mobile radio voice codecs in second generation systems, such as the Pan-European GSM, the American IS-54 and IS-95, as well as the Japanese systems. The corresponding bit rates are 13, 8, 9.6 and 6.7 kb/s, respectively. As an example, the proposed 11.36 kb/s prototype Codec 1 and the 11 kb/s Codec 2 are embedded in the adaptively reconfigurable wireless videophone Systems 1–4 featured in Table IV and their video quality, bit rate, robustness, and complexity issues are investigated. Coherent reconfigurable 16 or four-level pilot symbol assisted quadrature amplitude modulation (PSAQAM) is used and the system's robustness is improved by a combination of diversity and automatic repeat request (ARQ) techniques. When using a bandwidth of 200 kHz, as in the Pan-European GSM mobile radio system, the number of videophone users supported varies between three and 16, while the minimum required channel signal to noise ratio over Gaussian and Rayleigh channels is in excess of 6 and 8 dB, respectively, assuming a noise-limited, rather than interference-limited scenario. The salient system features are summarized in Table IV.**

I. Wireless Video Telephony

IN parallel with the proliferation of wireless multimedia services [1], a plethora of video codec schemes have been proposed for various applications [4], [5], but the perhaps most significant advances in the field are hallmarked by the MPEG-4 initiative [16]. A major feature topic of the European Community's recently initiated Fourth Framework Programme [17] on Advanced Communications Technologies and Services (ACTS), is video communications over a range of wireless and fixed links.

This treatise is focused on the design and performance evaluation of wireless video telephone systems, suitable for the robust transmission of quarter common intermediate format (QCIF) sequences over conventional mobile radio links, such as the Pan-European GSM system [18], the American IS-54 and IS-95 systems, as well as the Japanese digital cellular (JDC) system.[1] In contrast to existing standard codecs, such as the CCITT H261, H26P schemes and the MPEG-1 [2], MPEG-2 [3], and MPEG-4 [16] arrangements, our proposed video codec's fixed, but arbitrarily programmable bitrate facilitates its employment also in future intelligent systems, which are likely to vary their bit rate in response to various propagation and teletraffic conditions. The proposed QT codec is then embedded into a reconfigurable linear transceiver and its performance is studied in a multiuser scenario over nondispersive narrowband Rayleigh fading and additive white Gaussian noise (AWGN) channels [19]. Our experimental channel conditions are described by a propagation frequency of 1.9 GHz, vehicular speed of 30 mi/h and a signalling rate of 144 kBd, requiring a bandwidth of 200 kHz, as in the case of the Pan-European GSM system [18]. The quadtree- (QT)-coded video stream was transmitted using a reconfigurable transceiver, which has a more noise- and interference-resilient but less bandwidth efficient four-level quadrature amplitude modulation (4QAM) mode of operation or can halve its Bd-rate requirement, if better channel conditions are maintained by invoking the more bandwidth-efficient four-bit/symbol 16QAM mode [21].

The paper is structured as follows. Section II details the design of the 11.36 kb/s QT codec and its parametric enhancement, which is followed by a brief discussion on transmission issues in Section III, including modulation, forward error correction (FEC) coding and automatic repeat request (ARQ) techniques. The system's performance is characterized in Section IV, while Section V offers some conclusions.

II. Quadtree Codec Design

A. QT Codec Outline

The block diagram of the QT codec processing 176 × 144 pixels QCIF images scanned at 10 frames/s is shown in Fig. 1. For videophony over conventional mobile radio speech channels, such as the Pan-European GSM system [18], fixed-rate video codecs are required. As it is seen at the top of the codec's schematic, the intra-frame coding mode is invoked at the commencement of communications, during which a low-resolution initial image is transmitted to the decoder in order assist in its start-up phase, as will be described shortly. In the motion prediction (MP) block of Fig. 1, the QCIF frame is first segmented into small, for example 8 × 8-pixel, perfectly tiling blocks. Then each block is slid over a certain motion-velocity and frame-scanning rate dependent search area of the previous reconstructed frame and it is estimated by finding the position of highest correlation, which location each block was deemed to have originated from due to motion translation. The corresponding coordinates referred to as motion vectors (MV)

Manuscript received June 29, 1995; revised November 27, 1995. This paper was recommended by Associate Editor H. Gharavi. This work was supported by the EPSRC, UK under research contract GR/J46845.

The authors are with the Department of Electrical and Computer Science, University of Southampton, SO17 1BJ, UK.

Publisher Item Identifier S 1051-8215(96)03013-3.

[1] This work is supported by a demonstration package down-loadable from http://www-mobile.ecs.soton.ac.uk portraying various video sequences scanned at 10 frames/s and encoded at fixed bit rates of 6.7, 8, 9.6, 11.4, and 13 kb/s. Exposition of the paper is greatly augmented by displaying the corresponding video sequences.

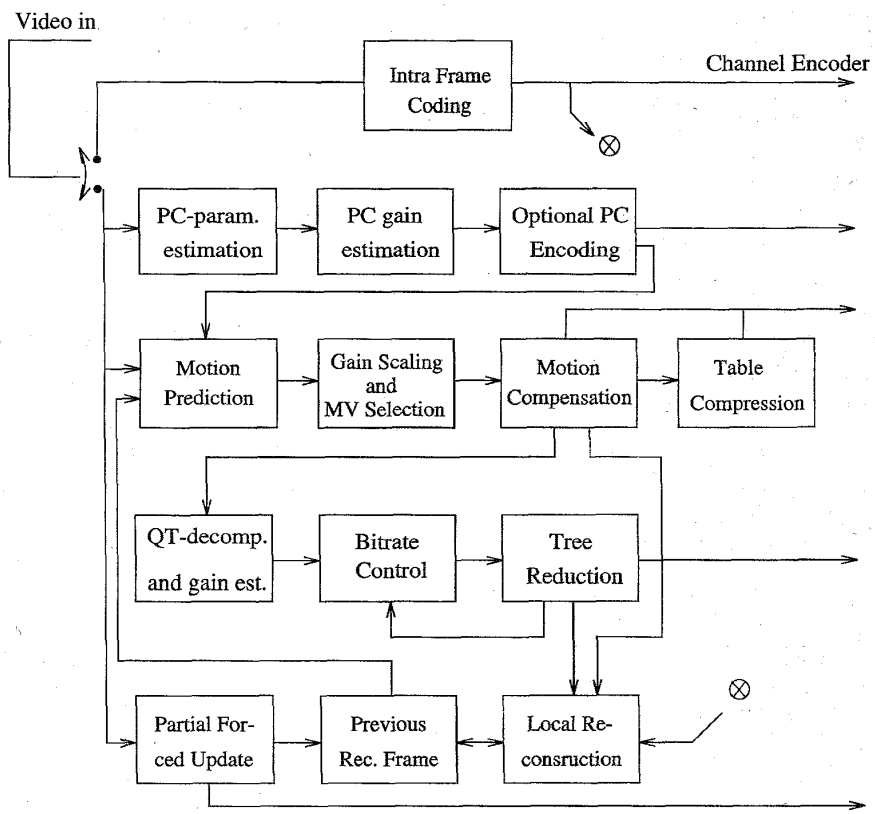


Fig. 1. QT codec schematic.

are then used in the motion compensation (MC) process of Fig. 1 to appropriately position each incoming block, which is then subtracted from the previously reconstructed frame in order to generate the so-called motion compensated prediction (MCP) residual. The MCP residual can then be modeled by a range of techniques, which will be addressed at a later stage. In the reconstruction process, the MV's assist at both the local and the distant decoder in an inverse fashion in order to appropriately update the reconstruction frame buffers.

Due to transmission errors, the encoder's and decoder's previous reconstructed frame buffers may become misaligned, which leads to prolonged artifacts at the output of the decoder. This effect can be remedied using partial forced updates (PFU) in order to identically replenish the encoder's and decoder's previous reconstructed frame buffers. In our proposed codecs, we were constrained to a simple PFU technique using the down-scaled and partially overlaid four-bit encoded 8×8-bit block averages, which was applied only to a fraction of the blocks in each frame due to the tight bit rate budget available. Although the PFU process partially overwrites the previous reconstructed buffers at both the local and remote decoder by the above mentioned very crude image estimate, therefore slightly impairing the codec's error-free performance, in case of high channel bit error rates (BER), it has an error mitigating effect by gradually replenishing both buffers' contents.

The number of 8×8-pixel PFU blocks per QCIF frame depends on the target bit rate and is automatically determined by the proposed reconfigurable codec. Consequently, the frequency of partially forced updating a certain block is also bit rate dependent, although higher prevailing BER values would require more frequent updates, irrespective of the bit rate budget. In our 11.36 kb/s QT videophone codec 20 randomly scattered 8×8 blocks out of the 396 blocks per frame are updated, which requires 80 b/frame. This implies that the PFU frequency of each specific block is about 1/(2 s), which is equivalent to updating the same block about every two seconds or 20 frames. During PFU, both the local and remote reconstructed buffers' contents are scaled by 0.7 and the 0.3-scaled four-bit quantized block averages are superimposed, allowing a nondestructive gradual replenishment to take place.

The MC scheme determines a four-bit MV for each of the 8×8 blocks within a search window of $4 \times 4 = 16$ pixels using full-search. The potential gain of MC is assessed in terms of MCP energy reduction and the gains in the subjectively important eye and mouth region may be augmented by a factor of two. A bit-rate dependent number of "motion-active" blocks is then subjected to full motion compensation, while for the "motion-passive" blocks, frame differencing is employed. Each of the 396 blocks would require a nine-bit identifier, leading to a total of $9 + 4 = 13$ bits per active vector. Typical motion activity rates of around 60 MV's out of 396 would exhaust most of the available bit rate budget of about 1000 bits per QCIF frame. In order to accommodate around 60 active MV's within a budget of 500 bits, we assign a one-

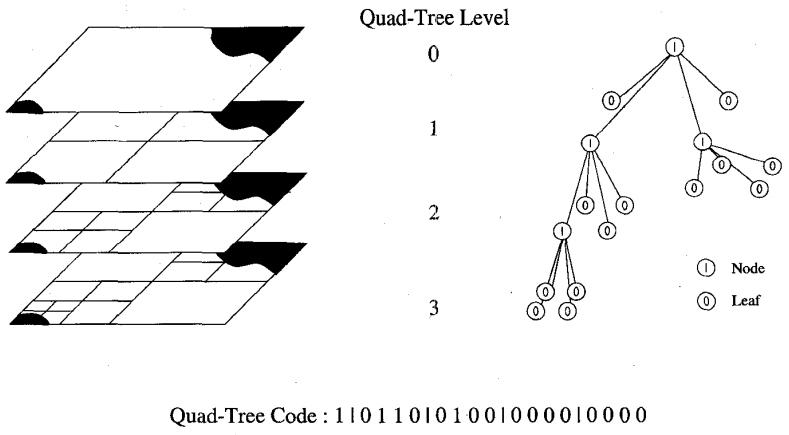


Fig. 2. Regular decomposition example and the corresponding quadtree.

bit motion-activity flag for each vector and hence create an MV activity table of 396 bits. This motion activity table can be compressed by about a factor of two using the run-length based "Table Compression" algorithm of Fig. 1, when aiming for a target bit rate budget of 1136 b/frame. The final bit allocation scheme of our prototype codec will be revisited in Table V after unraveling further details of the QT codec of Fig. 1.

The MCP residual frame can then be represented using a range of techniques, including subband coding (SBC) [6], [7], wavelet coding [8], discrete cosine transformation (DCT) [22], [2], [3], vector quantization (VQ) [11], [12], or QT [9], [10], [14] coding. In the proposed system we opted for the latter, which will be described in depth in the forthcoming sections. Observe, however, at the top part of Fig. 1 that an optional so-called model-based parametric coding (PC) scheme can also be invoked in the subjectively most important eye and lip regions of the image. Although this parametric codec was invoked in our system, the issues of PC are beyond the scope of this treatise.

B. Quadtree Decomposition

In QT coding [9], [10], [14], the MCP residual is described by the help of variable size sectors characterized by similar features, for example, by similar grey levels. Pursuing the so-called top-down QT decomposition approach of Fig. 2, let us derive the variable-length QT-code given at the bottom of the figure. The 176 $\times$ 144-pixel QCIF MCP residual frame constitutes a so-called node in the QT. After splitting, this node gives rise to four further nodes, which are classified on the basis of the "similarity criterion." Specifically, if all the pixels at this level of the QT differ from the mean m by less than the threshold σ, then they are considered to be a so-called "leaf node" in the QT. Hence, they do not have to be subjected to further "similarity tests," they can be represented simply by the mean value m. In our example a four-level decomposition was used (0–3), but the number of levels and/or the similarity threshold σ can be arbitrarily adjusted in order to achieve the required image quality and/or bit rate target as well as to emphasize the subjectively most important eye-lip region without explicitly informing the decoder about it.

If, however, the pixels constituting the current node to be classified differ by more than the threshold σ, the pixels forming the node cannot be adequately represented by their mean m and thus they must be further split until the threshold condition is met. This repetitive splitting process can be continued until there are no more nodes to split, since all the leaf nodes satisfy the threshold criterion. Consequently, the QT structure describes the contours of similar grey levels in the frame difference signal. The derivation of the QT-code now becomes explicit from Fig. 2, where each parent node is flagged with a binary one classifier, while the leaf nodes are denoted by a binary zero classifier and the flags are read from top to bottom and left to right. Observe in the example of Fig. 2 that the location and size of 13 different blocks can be encoded using a total of 17 bits.

The typical segmentation of a frame is exemplified in Fig. 3, where the QT structure is portrayed at the left-hand side of the figure. In the center of Fig. 3, the corresponding overlaid MCP residual frame is displayed, while the original video frame is portrayed at the right-hand side of the figure. In the eye and lip regions, a more stringent similarity match was required than in the background, which led to a finer QT decomposition. It is worth pointing out that in case of high image quality requirements, the simple block average representation of the QT decomposed blocks can be substituted by more sophisticated techniques, such as VQ, DCT, or subband (SB) decomposed representations. Naturally, increasing the number of hierarchical levels in the QT decomposition leads to blocks of different sizes, and applying VQ or DCT to blocks of different sizes increases the codec's complexity. Hence, the employment of these schemes in case of more than two to three hierarchy levels becomes impractical. Having reviewed the principles of QT decomposition, let us now focus our attention on the design of our candidate codecs.

C. Quadtree Intensity Match

With low implementational complexity in mind, we contrived two candidate codecs. In the first one we used the above introduced zero-order mean- or average-based decomposition, while in the second one we attempted to model the luminance intensity profile over the block by a first-order linear approx-

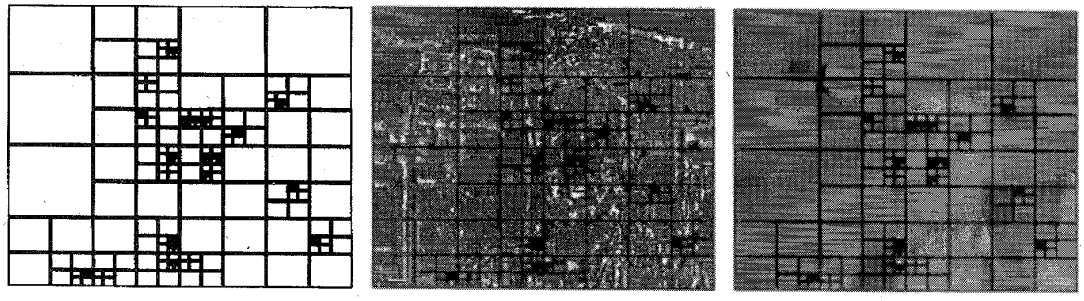


Fig. 3. Quadtree segmentation example with and without overlaid MCP residual and original video frame.

TABLE I
PSNR PERFORMANCE COMPARISON FOR ZERO- AND FIRST-ORDER QT MODELS AT VARIOUS CONSTANT BIT RATES

Level	0	1	2	3	4	5	6
Block Size	176× 144	88× 72	44× 36	22 × 18	11× 9	5/6 × 4/5	2/3 × 2/3
Max. Range +/-	0.22/0.36	0.57/0.79	2.89/2.49	8.48/7.74	20.59/20.69	57.59/50.39	97.26/86.62

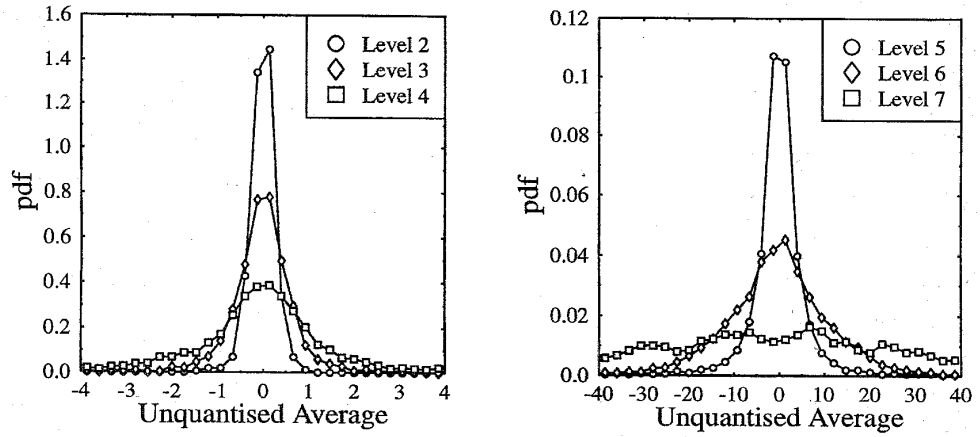


Fig. 4. Probability density function of the block averages at Levels 2 to 7.

imation corresponding to luminance plane sloping in both x and y directions.

Zero-Order Intensity Match: The statistical evaluation of the probability density functions (PDF) of the average values m of the variable sized MCP blocks portrayed in Fig. 4 revealed that for various block sizes, quantizers having different mean values and variances were required. As the above mentioned mean-value PDF's revealed, the mean of the MCP blocks toward the top of the QT, namely at QT Levels 2–4, which cover a large area, is more likely to be close to zero than the mean of smaller blocks, which exhibits itself in a more highly peaked and hence less spread PDF. Observe that for clarity of visualization, the PDF scales of Levels 2–4 and 5–7 are about an order of magnitude different. The mean of the smallest blocks at QT Levels 5–7 tends to fluctuate over a wide range, yielding a near-uniform PDF for Level 7. Table I summarizes the intervals over which the block means fluctuate as the block size is varied. Therefore, it was necessary to design separate quantizers for each QT hierarchy level.

With this objective in mind, a two-stage quantizer training was devised. First the unquantized mean values for each QT-decomposed block size were recorded using a training sequence in order to derive an initial set of quantizers. During the second, true training stage, this initial set of quantizers was then used tentatively in the QT codec's operation in order to record future unquantized block averages generated by the codec operated at the chosen limited, constant bit rate, namely at 1000 b/frame. This two-stage approach was necessary to obtain realistic training data for the Max–Lloyd training of the final quantizers.

In order to achieve the best codec performance under the constraint of generating 1000 bits per frame, we derived codebooks for a range of different number of reconstruction levels. Fig. 5 characterizes the codec's performance for various quantizers ranging from two to 64 level schemes. Observe in the figure that the two-level and four-level quantizers were found to have the best performance. This was due to the fact that the 10 kb/s bit rate limit severely restricted the number of hierarchy levels in the QT decomposition process, when more bits were allocated to quantize the block averages. Hence, the codec failed to adequately decompose the inhomogeneous regions due to allocating an excessive number of bits to the quantization of their mean values.

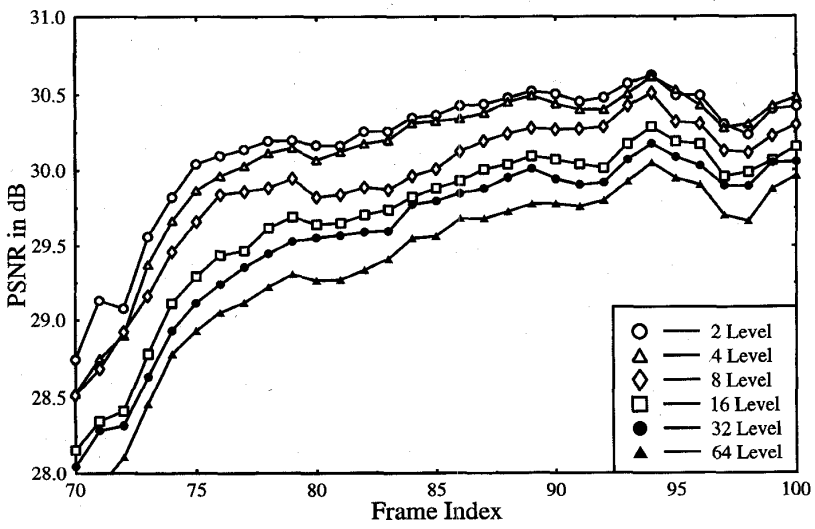


Fig. 5. Performance for various quantizers at a constant bit rate.

First-Order Intensity Match: Following our endeavors using the zero-order model within the QT decomposed subblocks, we then embarked on studying the performance of this technique using the bidirectionally sloping linear luminance surface of: $b(x,y) = a_0 + a_{x1}x + a_{y1}y$, where for image reconstruction, the quantized values of a_{x1} and a_{y1} must be known. The squared error of this linear approximation is given by (1). The constants a_0, a_{x1}, and a_{y1} are determined by setting the partial derivative of the error term

$$e = \sum_{x=1}^{b}\sum_{y=1}^{b}(b(x,y) - (a_0 + a_{x1}x + a_{y1}y))^2 \qquad (1)$$

with respect to all three variables to zero and solving the resulting 3-D problem. In order to assess the performance of this scheme, Max–Lloyd quantizers were designed for each of the constants and the peak signal-to-noise ratio (PSNR) performance of a range of quantizers using different numbers of quantization levels was tested. We found that the increased number of quantization bits required for the three different coefficient quantizers was too high to facilitate a sufficiently fine QT-based MCP residual decomposition. Hence, the PSNR performance of the first order scheme became inferior to that of the zero-order model under the constraint of a fixed bit rate. Specifically, Table II reveals that the PSNR performance of the first-order model is at least 0.5 dB worse than that of the simple zero-order model. Consequently, for our further experiments, the latter scheme was preferred.

TABLE II
PSNR PERFORMANCE COMPARISON FOR ZERO- AND FIRST-ORDER QT MODELS AT VARIOUS CONSTANT BIT RATES

Approximation	8 kb/s	10 kb/s	12 kb/s
Zero-order PSNR (dB)	27.10	28.14	28.46
First-order PSNR (dB)	25.92	27.65	27.82

D. QT Decomposition Algorithmic Issues

The crucial element of any QT codec is the segmentation algorithm controlled by the similarity threshold σ. Most codecs presented in the recent literature by Strobach [9], Arnold *et al.* [13], as well as by Shusterman and Feder [14] invoked the previously outlined thresholding process. This principle results in an approximately time-invariant video quality and a fluctuating, time-variant bit rate. This bit rate fluctuation can be smoothed applying adaptive buffer feedback techniques at the cost of increased video delay or by controlling the similarity threshold σ [9].

Furthermore, in their impressive contribution, Shusterman and Feder [14] proposed techniques for improving the rate-distortion (RD) performance of still-image QT codecs by attempting to optimize the absolute difference similarity threshold and bit allocation. Explicitly, the authors suggested using a different absolute difference similarity threshold at different QT resolution levels, which was determined by minimizing a mean-squared error (mse) upper bound, rather than the mse itself, since there is no closed form expression for the mse as a function of the absolute difference similarity threshold. Their conclusion was to opt for halving the similarity threshold at each hierarchy level, which is plausible on the basis of considering the reduced area and mse contribution of lower QT levels. They used Max–Lloyd quantization and an approximately 10% transmission overhead for the mean and variance of all levels, while maintaining a PSNR value of 28.91 dB for the Lena still picture at a coding rate of 0.5 b/pixel.

In contrast, since our proposed videophone system was designed for constant-rate mobile radio speech channels, we opted for a constant, but arbitrarily programmable low bit rate motion-picture coding scheme, where the coding rate is lower, around 1000 b/frame/25 344 pixels/frame $\approx$ 0.05 b/pixel, while maintaining a similar PSNR for the well-known Miss America (MA) sequence. A further requirement in very low bit rate video communications is gain-cost control, which is difficult to achieve in the framework of QT schemes, since the gain of a potential decomposition becomes known only after the decomposition took place. For these reasons, we tested the

performance of a range of algorithms and arrived at the following novel gain-cost controlled bit allocation algorithm, which refrains from using the thresholding technique of Shusterman and Feder and is summarized in Algorithm 1.

Algorithm 1: This algorithm adaptively adjusts the required QT resolution, the number of QT description bits, and the number of encoding bits required in order to arrive at the target bit rate.

```
1. Develop the full tree from minimum to
maximum number of QT levels (e.g., 2-7).
2. Determine the mse gains associated
with all decomposition steps for the
full QT.
3. Determine the average decomposition
gain over the full set of leaves.
4. If the potentially required number
of coding bits is more than twice
the target number of bits for the
frame, then delete all leaves having
less than average gains and repeat
Step 3.
5. Otherwise, delete leaves on an
individual basis, starting with the
lowest gain leaf, until the required
number of bits is attained.
```

In our proposed approach, the codec develops the QT structure down to a given maximum number of decomposition levels and then determines the gain of each decomposition step, as summarized in Steps 1 and 2 of Algorithm 1. Exposition of the algorithm is further aided by referring to Fig. 2. The decomposition gain is defined as the difference between the mean squared video reconstruction error of the "parent block" and the total mse contribution associated with the sum of its four "child blocks." The question arose whether the potential gain due to the current additional decomposition step should be weighted by the actual area of the blocks under consideration. Toward the bottom of the QT, the subblocks represent small image sections, while toward the top, larger areas. Our simulations showed that for most of the cases, such a weighting of the gains was disadvantageous. This somewhat surprising result was an indirect consequence of the constant bit rate requirement, since the adaptive codec often found it more advantageous to resolve some of the fine image details at an early stage, if the bit rate budget allowed it, and then save bits during future frames over the area in question.

Returning to Algorithm 1, the purpose of Steps 3 and 4 is to introduce a bit rate-adaptive, computationally efficient way of pruning the QT to the required resolution. This allows us to incorporate an element of cost-gain quantized coding, while arriving at the required target bit rate without many times tentatively decomposing the image in various ways in an attempt to find the optimum fixed bit allocation scheme. The algorithm typically encountered four to five such fast QT pruning recursions, before branching out to Step 5, which facilitated a slower converging fine-tuning phase during the bit allocation optimization.

Clearly, Algorithm 1 allowed us to eliminate the specific child-blocks or leaves from the tree that resulted in the lowest decomposition gains. During this QT pruning process, the elimination of leaves converted some of the nodes to leaves, which were then considered for potential elimination during future coding steps; therefore, the list of leaves associated with the lowest decomposition gains had to be updated before each QT pruning step. During the fast pruning phase constituted by Steps 3 and 4 in a computationally efficient, but suboptimum approach, we determined the average gain of the entire set of leaves, and rather than deleting the leaves associated with the lowest decomposition gain one-by-one, all leaves having a gain lower than the average gain were deleted in a single step. The slower one-by-one pruning constituted by Step 5 was then invoked before concluding the bit allocation in order to fine-tune the number of bits allocated. This suboptimum deletion process was repeated until the tree was pruned to the required size and the targeted number of coding bits was allocated.

In an attempt to explore the eventual residual redundancy within the QT, we examined the tree structure. The philosophy of the following procedure was to fix the minimum and maximum tree depth. The advantage of such a restriction is twofold, as both the number of bits required for the QT and the complexity can be reduced. Under this QT structure limitation, bits can be saved both at the top and at the bottom of the tree. If a certain minimum depth of the tree is inherently assumed, the tree information of the top levels becomes *a priori* knowledge. The same happens with regard to the bottom level, if the tree depth is limited. In that case, all decomposed subblocks at the last level are children, implying that no further splitting is allowed to take place. There is no need to transmit this *a priori* information. Hence, the decomposition complexity is limited, since the tree needs to be developed only for the tree branches, which do not constitute *a priori* knowledge.

In Table III we listed both the average and maximum number of nodes associated with each tree level. Observe from the table that $15.3/16 = 95.6\%$ of the time, the QT depth exceeds three levels. Hence, it is advantageous to restrict the minimum tree depth to three without sacrificing coding efficiency. Note from Table III, furthermore, that only about $52.9/16\ 384 = 0.32\%$ of the potentially possible Level 7 leaves is ever decomposed. Hence, if we restrict the maximum tree depth to six, the codec's complexity is reduced, which is associated with a concomitant average PSNR reduction from 33.78 dB to 32.79 dB for the first 100 "Miss America" frames.[2] As seen from Table III, limiting the QT depth to six levels increases the number of Level 6 leaves by 52.9, which is the number of relegated Level 7 leaves. These now have to be allocated additional reconstruction level coding bits. On the other hand, due to this restriction, the $(72.6 + 52.9) = 125.5$ Level 6 leaves do not have to be specifically flagged as leaves in the variable-length QT code, which results in bit savings. When using a four-level quantizer and a bit rate budget of 1000 b/frame, on the basis of simple logic, one would expect a similar performance with and without Level 7 decomposition, since we could save 125.5 QT description bits, which then could be invested to quantize $\mathrm{INT}[125.5/2] = 62$

[2]Various video sequences encoded at 6.7, 8, 9.6, 11.36 and 13 kb/s can be viewed under the following WWW address: http://www-mobile.ecs.soton.ac.uk.

TABLE III
AVERAGE TREE NODES AND LEAVES PER HIERARCHICAL LEVEL WITHIN THE TREE AT 1200 BPF

Level	0	1	2	3	4	5	6	7
Average nodes/leaves	1	4	15.3	28.4	47.6	66.9	72.6	52.9
Max. poss. nodes/leaves	1	4	16	64	256	1024	4096	16384

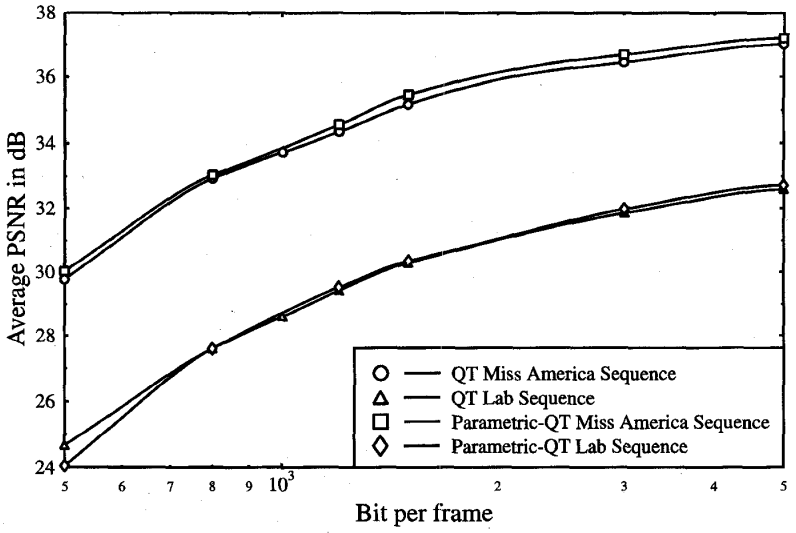


Fig. 6. PSNR versus bit rate performance of the proposed QT codec with and without PC enhancement.

newly created reconstruction levels. Nonetheless, the 1 dB PSNR reduction encountered due to removing Level 7 from the QT suggested that the adaptive codec reacted differently, although a complexity reduction was achieved. Hence, if video quality was at premium, the seventh QT level was necessary to develop important fine details in the frame.

The achievable PSNR performance of the proposed QT codec at various bit rates is portrayed in Fig. 6 in case of the Miss America sequence and a higher activity locally recorded clip, which we referred to as the "Lab Sequence." This figure characterizes the expected codec performance in an adaptive communicator of the near future, where improved video quality is achieved, for example, at twice the bit rate using the double bit rate but more vulnerable 16-level quadrature amplitude modulation (16QAM) modem [21], while maintaining the bandwidth requirement of the halved bit rate but more robust 4QAM scheme. These issues will addressed in more depth during our further discussions. Observe in the figure that for the lower activity Miss America sequence, a consistently better PSNR was achieved. Notice furthermore that the PSNR improvement of the model-based PC enhancement of Fig. 1 is also portrayed in the figure, although its objective PSNR effects appear mitigated by the relatively small area of the eyes and lips, where it is applied. Its subjective effects are more pronounced. In order to comply with our initial goal of contriving a QT codec, which can be accommodated by a conventional mobile radio voice channel, we opted for an 11.36 kb/s QT codec in our videophone system. Having described the proposed QT codec, let us now focus our attention on transmission issues.

III. TRANSMISSION ASPECTS

A. Channel Coding and Modulation

Trellis coded modulation (TCM) or block coded modulation (BCM) [23], [21] use modulation constellation expansion in order to accommodate the parity bits required for error correction coding and they achieve usually higher coding gain than isolated coding and modulation. However, the bits generated by our QT video codec have unequal bit sensitivities and therefore unequal protection channel coding [7], [28] must be invoked. In the spirit of [7] and [24], here we employ source-sensitivity matched unequal protection joint source/channel coding and modulation schemes. The QT-coded video bits are mapped in two different protection classes, namely, class one and two. Both streams are binary Bose–Chaudhuri–Hochquenghem (BCH) coded [20] and bursted out using 4- or 16-QAM [21] and time division multiple access (TDMA). In order for the transmitted signal to fit into a bandwidth of 200 kHz while using an excess bandwidth of 38.8% corresponding to a Nyquist roll-off factor of 0.388, the signalling rate was limited to 144 kBd.

Differentially coded noncoherent modems have lower implementational complexity than coherent modems [21], but the corresponding rotationally symmetric noncoherent 16QAM phasor constellation has a weakness in terms of "error-doubling" due to differential coding. Hence, the coherent rectangular 16QAM constellation requires typically about 3 dB lower channel SNR. This property prompted us to opt for the latter constellation. Coherent detection was achieved using pilot symbol assisted modulation [21] (PSAM), where

TABLE IV
SUMMARY OF SYSTEM FEATURES

Feature	System 1	System 2	System 3	System 4
Video Codec	Codec 1	Codec 2	Codec 1	Codec 2
Video rate (kbps)	11.36	11	11.36	11
Frame Rate (fr/s)	10	10	10	10
C1 FEC	BCH(127,71,9)	BCH(127,50,13)	BCH(127,71,9)	BCH(127,50,13)
C2 FEC	BCH(127,71,9)	BCH(127,50,13)	BCH(127,71,9)	BCH(127,50,13)
Header FEC	BCH(127,50,13)	BCH(127,50,13)	BCH(127,50,13)	BCH(127,50,13)
FEC-coded Rate (kbps)	20.32	27.94	20.32	27.94
Modem	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM
ARQ	No	No	Yes	Yes
User Signal. Rate (kBd)	18 or 9	12.21 or 24.75	18 or 9	12.21 or 24.75
System Signal. Rate (kBd)	144	144	144	144
System Bandwidth (kHz)	200	200	200	200
No. of Users	8 or 16	5 or 11	6 or 14	3 or 9
Eff. User Bandwidth (kHz)	25 or 12.5	40 or 18.2	33.3 or 14.3	66.7 or 22.2
Min. AWGN SNR (dB) 4/16QAM	7.5/13	7.5/12	6/12	6/11
Min. Rayleigh SNR (dB) 4/16QAM	20/20	15/18	8/14	8/14

"channel sounding" symbols with known phase and magnitude are inserted in the transmitted TDMA burst in order to inform the receiver about the channel's estimated momentary phase rotation and attenuation. The inverse complex-domain derotation and deattenuation can then be applied to the data symbols in order to remove the effects of the fading channel.

After extracting the pilot symbols from the received TDMA burst, a complex channel attenuation and phase rotation estimate must be derived for each received symbol using appropriate interpolation techniques. While previously proposed PSAM schemes used either a low-pass interpolation filter or an approximately Gaussian filter, we found that for a wide range of pilot spacing values, the best BER performance was achieved by a low-complexity polynomial interpolator [25]. In [7] and [21], we showed that the maximum minimum distance rectangular 16QAM constellation exhibits two independent two-bit subchannels. The lower integrity C2 subchannel showed a factor two to three times higher BER than the higher quality C1 subchannel. Although these integrity differences can be fine-tuned using different BCH codes in the subchannels, for our system it was found appropriate to maintain this integrity ratio for the class one and class two QT-coded bits.

B. System Architecture

Four different transceivers were contrived in order to explore the range of system design trade-offs. In System 1, we employed the $R = 71/127 \approx 0.56$-rate BCH$(127, 71, 9)$ code in both 16QAM subchannels. After BCH coding the 11.36 kb/s QT coded video sequence generated by Codec 1 the bit rate became $11.36 \times 127/71 = 20.32$ kb/s. These system features are tabulated in Table IV. In System 2, on the other hand, we opted for the stronger BCH$(127, 50, 13)$ code, which after FEC coding the 11 kb/s stream produced by Codec 2 yielded a bit rate of 27.94 kb/s. Since System 2 has a higher transmission rate, it will inevitably support a lower number of users than System 1. This comparison will allow us to decide in Section IV whether it is worthwhile in terms of increased system robustness using a stronger BCH code at the cost of reducing the number of users supported.

The Transmission Packets of System 1 are comprised by a Class One BCH$(127, 71, 9)$ codeword conveyed over the C1 16QAM subchannel and a Class Two BCH$(127, 71, 9)$ codeword transmitted over the C2 subchannel. A stronger BCH$(127, 50, 13)$ codeword is assigned to the packet header. The 381-bit packets are converted to 96 16QAM symbols and 11 pilot symbols are inserted with a pilot spacing of $P = 10$. Lastly, four ramp symbols are concatenated at both end of the packet in order to allow smooth power amplifier on/off ramping, which mitigates spectral spillage into adjacent frequency bands. Eight 111-symbol packets are required to transmit an entire 1136-bit image frame and hence the signalling rate becomes 111 symb/12.5 ms $\approx$ 9 kBd. This allows us to support 144 kBd/9 kBd =16 users, which is identical to the number of half-rate speech users accommodated by the GSM system [18]. The packet format of System 2 is identical to that of System 1, but 11 packets are required to transmit an entire 1100-bit/100 ms frame. Hence, the signalling rate becomes 11×111 symb/100 ms $= 12.21$ kBd. Then the number of users supported by System 2 at the 144 kBd TDMA rate is INT$[144 \text{ kBd}/12.21] = 11$, where INT[..] represents the integer part of [..].

When the channel signal-to-noise ratio (SNR) becomes insufficient for 16QAM communications, since, for example, the Portable Station (PS) moves away from the Base Station (BS), the BS reconfigures the transceivers as 4QAM schemes, which requires twice as many time slots but a lower SNR value. The 381-bit packets are now converted to 191 4QAM symbols and after inserting 20 pilot symbols and four ramp symbols, the packet-length becomes 225 symb/12.5 ms, yielding a signalling rate of 18 kBd. Hence the number of videophone users supported by System 1 is reduced to eight, as in the full-rate GSM speech channel. The signalling rate of the 4QAM mode of System 2 becomes 11×225 symb/100 ms $= 24.75$ kBd and the number of users is INT$[144/24.75] = 5$.

The system also supports mixed-mode operation, where the more error resilient 4QAM users must reserve two slots in each 12.5 ms TDMA frame, while roaming near the fringes of the cell. By contrast, in the central section of the cell, 16QAM users will only require one slot per frame in order to

TABLE V
BIT ALLOCATION FOR THE 11.36 kb/s CODEC 1

Parameter	MC	PFU	QT	Total
No. of bits	< 500	80	556	1136

maximize the number of users supported. Assuming an equal proportion of System 1 users in their 4QAM and 16QAM modes of operation, the average number of users per carrier becomes 12. The equivalent user bandwidth of the 4QAM PS's is 200 kHz/8 = 25 kHz, while that of the 16QAM users is 200 kHz/16 = 12.5 kHz. The characteristics of the whole range of our candidate systems are highlighted in Table IV.

Systems 3 and 4 are identical to Systems 1 and 2, respectively, except for the fact that the former schemes can invoke ARQ when the received video bits are erroneous. In the past, the employment of ARQ schemes was limited to data communications [27], where longer delays are tolerable than in interactive video telephony. In the proposed packet video system, however, there is a full duplex control link between the BS and PS, which can be used for message acknowledgment, while the short TDMA frame length ensures a low packet delay, hence ARQ can be invoked. Therefore, System 3 and 4 are expected to have a higher robustness than Systems 1 and 2, which dispense with ARQ-assistance. The penalty must be payed in terms of a reduced number of users supported, since the ARQ attempts occupy some of the time slots.

It is unrealistic to expect the system to operate at such low channel SNR values where the probability of packet corruption and hence the frequency of ARQ attempts is high, since then the number of users supported and the teletraffic carried would become very low. In order to compromise, the number of transmission attempts was limited to three, which required two earmarked slots for ARQ. Furthermore, in any frame, only one user was allowed to invoke ARQ, namely the specific user whose packet was first corrupted within the frame. For the remaining users, no ARQ was allowed in the current TDMA frame. An attractive feature of this ARQ scheme is that if three copies of the transmitted packet are received, majority logic decisions can be invoked on a bit-by-bit basis, should all three received packets be corrupted. The basic features of Systems 3 and 4 are also summarized in Table IV.

IV. SYSTEM PERFORMANCE

A. Bit Sensitivity Issues

For the proposed videophone system an 11.36 kb/s, 10 frames/s (fps) scanned codec was designed. The bit allocation scheme of the codec is summarized in Table V. A total of less than 500 bits are allocated to the motion compensation activity table and the motion vectors, 80 bits for PFU's, while the remaining 556 bits are dynamically assigned to variable-length QT coding using the previously highlighted QT pruning as well as to the four-level QT-block mean quantization.

An often employed technique of quantifying the error sensitivity of source coded bits is to consistently invert a particular bit in every frame and evaluate the PSNR degradation inflicted. The deficiency of this technique is that it does not take adequate account of the different error propagation properties of different bits in consecutive frames through the reconstructed frame buffer. Here we take account of the error propagation effects by finding the average PSNR degradation due to a single bit error both in the frame in which the error occurs as well as in consecutive frames and integrate, i.e., sum, the PSNR degradations over the frames where degradations occur, as we proposed in [26]. In this context it becomes clear that without the error mitigating effects of the PFU technique employed, the transmission errors would have prolonged effects due to the misaligned reconstructed frame buffers of the encoder and decoder.

The error sensitivity of the PFU, the MV, and the reconstruction levels for the QT-decomposition bits evaluated according to the previously outlined principle is portrayed in Fig. 7. Note that the sensitivity of the four PFU bits and that of the $2 + 2 = 4$ MV bits follows the natural trend from the most significant bits (MSB) toward the least significant bits (LSB). As Fig. 7 demonstrates, the bit sensitivity of the QT reconstruction level bits increases toward smaller block sizes or deeper QT levels. This is due to the fact that the average luminance of smaller blocks has more spread distribution, as we have seen in Fig. 4 and hence, their quantization levels are more sparsely allocated, which results in an increased vulnerability, manifesting itself in terms of higher objective and subjective video degradation when corrupted. The variable-length quadtree code itself can be interpreted as run-length encoded information, since a corrupted bit affects the entire tree decoding process. Therefore, we refrain from quantifying their sensitivities and in case of their corruption, we rely on retransmission-based error control [27].

In order to achieve finely-graded sensitivity-matched FEC protection, rate-compatible punctured convolutional (RCPC) codes have been suggested in the past by Cox *et al.* [28]. However, in [26], the performance of a quad-class scheme was found virtually indistinguishable from that of a twin-class scheme. Hence, here we also opted for a twin-class scheme using the BCH(127, 71, 9) or the BCH(127, 50, 13) codecs in both the higher integrity C1 and lower integrity C2 16QAM subchannels for the transmission of the more sensitive and less sensitive video bits, respectively. In case of Codec 1, there are 1136/2 = 568 class one and 568 class two bits, while for Codec 2, 550 and 550, respectively. Although the variable-length QT code bits nearly fill the capacity of the C1 subchannel, up to 68 of the high-sensitivity PFU and MV MSB's of Fig. 7 can be directed to the higher integrity C1 subchannel.

B. Video System Performance

Let us now focus our attention on the overall video system performance evaluated in terms of the PSNR versus channel SNR curves depicted in Figs. 8–11 for Systems 1–4. In all figures, six performance curves are displayed, three curves with ARQ using three transmissions (TX3) and three curves without ARQ, i.e., using a single transmission attempt (TX1). Both the 4QAM and 16QAM mode of operation of Systems 1–4 are featured over AWGN channels, Rayleigh-fading channels with diversity (RD), and with no diversity (RND). Perceptually unimpaired video performance was typically achieved by these systems when the PSNR degradation was less than about 1 dB.

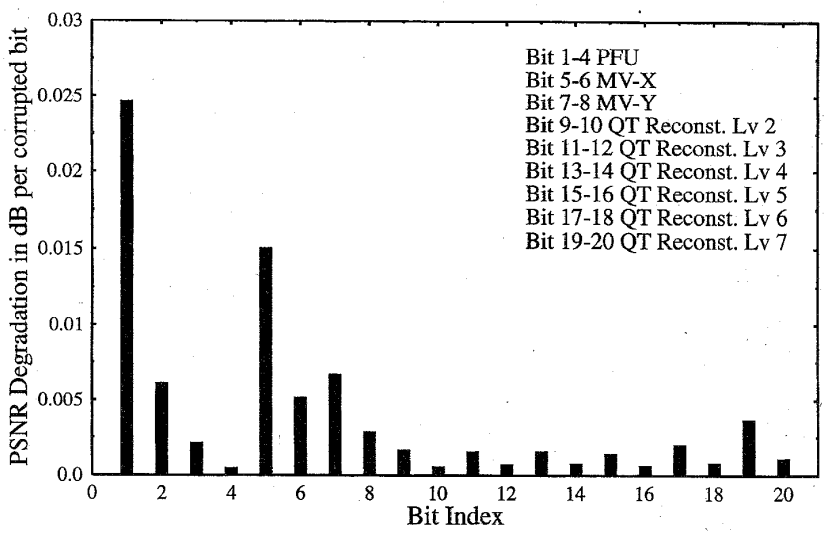


Fig. 7. Integrated bit sensitivities of the QT codec.

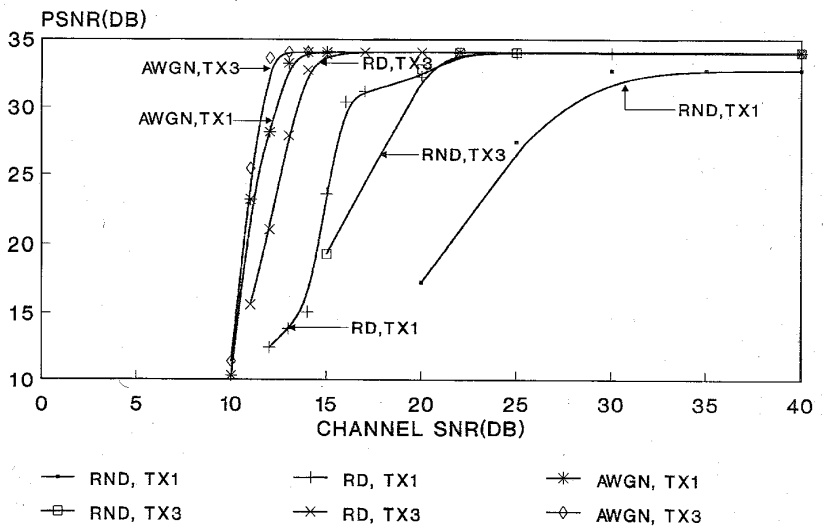


Fig. 8. PSNR versus channel SNR performance of the 16QAM mode of operation of Systems 1 and 3.

Therefore in Table IV we summarized the minimum required channel SNR values to ensure a PSNR degradation less than 1 dB, as the systems' operating channel SNR.

The full-scale exploration of the various system design trade-offs using Figs. 8–11 and Table IV is left to the reader for the sake of compactness. We will restrict our discourse to the description of some of the prevailing trends, which, in general, follow our expectations, rather than comparing each system to its counterparts. As anticipated, the most robust performance was achieved by all systems over AWGN channels, typically requiring a channel SNR of about 11–12 dB, when using the 16QAM mode of operation. In the 4QAM mode, the minimum required channel SNR was around 6–7 dB. The benefits of using ARQ over AWGN channels were limited to a channel SNR reduction of about 1 dB in almost all modes of operation, while reducing the number of users supported by two. This was a consequence of the stationary, time-invariant channel statistics, which inflicted similar channel impairments during the retransmission attempts. In contrast, in case of Rayleigh channels, the error-free reception probability of a packet was typically substantially improved during the ARQ attempts, effectively experiencing time-diversity. Hence, during re-transmissions, the PS often emerged from the fade by the time ARQ was invoked. The employment of the stronger BCH$(127, 50, 13)$ code failed to substantially reduce the required channel SNR's over Gaussian channels.

Over Rayleigh channels, the effect of diversity was extremely powerful in case of all systems, typically reducing the minimum channel SNR requirements by about 10–12 dB in the 4QAM mode and even more in the 16QAM modes of operation in case of both the weaker and stronger BCH codes. This was a ramification of removing the residual BER of the

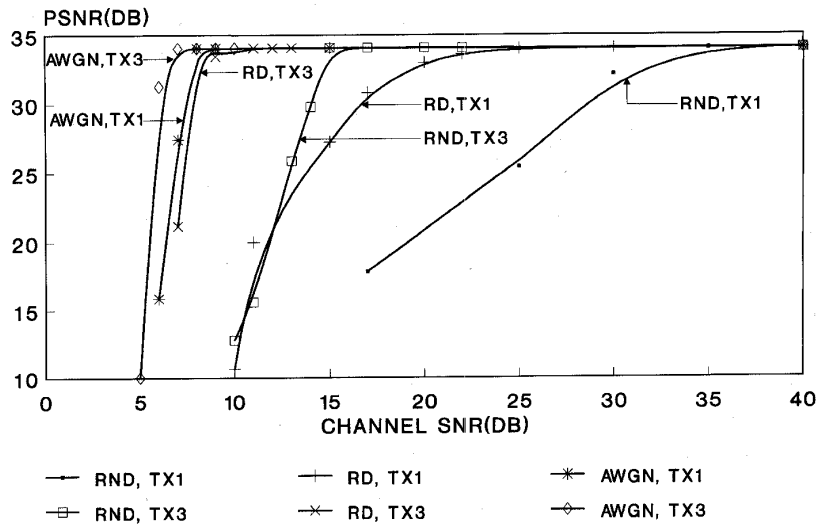


Fig. 9. PSNR versus channel SNR performance of the 4QAM mode of operation of Systems 1 and 3.

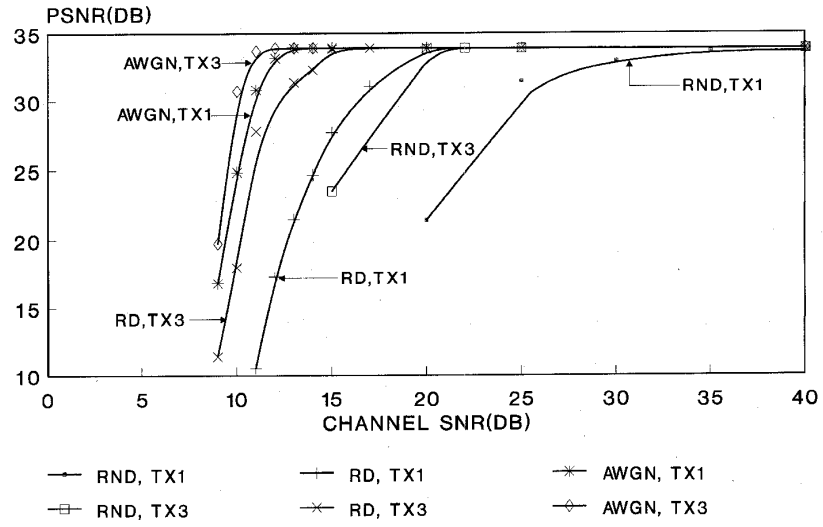


Fig. 10. PSNR versus channel SNR performance of the 16QAM mode of operation of Systems 2 and 4.

modems. In the 4QAM mode without diversity, the stronger BCH$(127, 50, 13)$ code reduced the channel SNR requirement by about 8 dB and with diversity, by 5 dB, when compared to the systems using the weaker BCH code. On the same note, in the 16QAM mode without diversity the BCH$(127, 50, 13)$ code managed to eradicate the modem's residual BER, but with diversity, its supremacy over the higher user-capacity, lower complexity BCH$(127, 71, 9)$ coded system eroded substantially, to about 2 dB. This indicated that in System 1, the weaker BCH code struggled to remove the modem's residual BER, but with diversity assistance the employment of the stronger code became less important.

Similarly dramatic improvements were observed in case of using ARQ, in particular, when no diversity was employed or the weaker BCH code was used. Again, this was due to removing the modem's residual BER. Specifically, in the 4QAM mode without diversity and with the BCH$(127, 71, 9)$ code an SNR reduction of about 16 dB was possible due to ARQ, which reduced to around 12 dB in case of the stronger code. Remaining in the 4QAM mode but invoking diversity reduced the ARQ gain to about 7 dB for the inherently more robust BCH$(127, 50, 13)$ code, while the weaker code's ARQ gain in System 3 was only slightly less with diversity than without, amounting still to about 12 dB. Similar findings pertain to the 16QAM ARQ-gains, when comparing Systems 1 and 3 as well as 2 and 4, which were of the order of 10 dB without diversity and about 5 dB with diversity for both BCH codes.

Clearly, both diversity and ARQ's are mitigating the effects of fading and in broad terms they achieve similar robustness, but the teletraffic reduction due to the reserved time slots is a severe disadvantage of the ARQ assistance. Diversity, on the

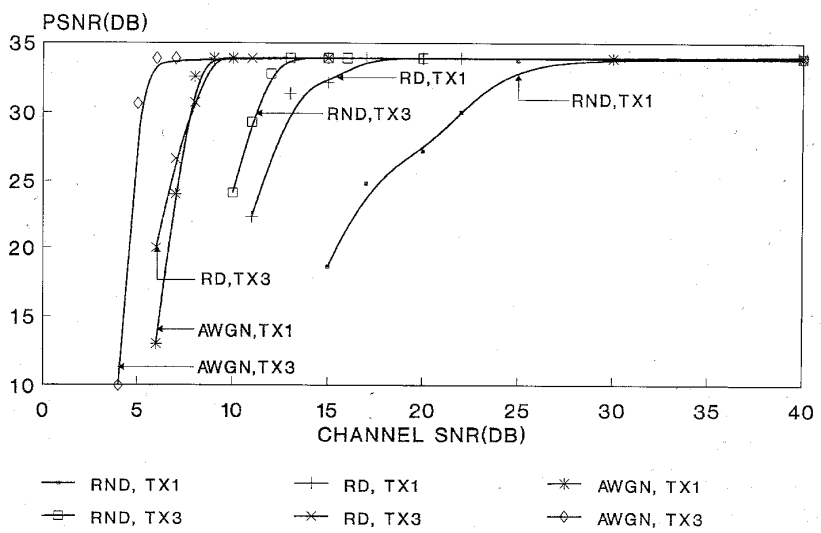


Fig. 11. PSNR versus channel SNR performance of the 4QAM mode of operation of Systems 2 and 4.

other hand, can readily be employed in small handsets operating at frequencies in excess of 1 GHz, as was demonstrated by the second generation Japanese digital voice handset, which is an attractive shirt-pocket sized consumer product. When using a combination of diversity and ARQ, near-Gaussian performance was maintained over fading wireless channels effectively by all four systems but, as evidenced by Table IV, at a concomitant low carried teletraffic.

Consequently, the number of users supported by the different systems and operating modes varies over a wide range, between three and 16, and the corresponding effective user bandwidth becomes 66.7 kHz and 12.5 kHz, respectively. The improved robustness of some of the ARQ-assisted schemes becomes less attractive in light of their significantly reduced teletraffic capacity. It is interesting to compare in robustness terms the less complex 4QAM mode of System 1 supporting eight users with the more complex 16QAM mode of System 4, which can accommodate a similar number of users, namely nine. The former necessitates a channel SNR of about 7 dB over AWGN channels, while the more complex latter scheme about 11 dB, indicating a clear preference in favor of the former. Interestingly, the situation is reversed over the diversity assisted Rayleigh channel, requiring SNR values of 20 and 14 dB, respectively. A range of further interesting system design aspects can be inferred from Figs. 8–11 and Table IV in terms system complexity, robustness, and carried teletraffic.

V. Summary and Conclusion

A range of reconfigurable QT-based, parametrically enhanced wireless videophone schemes have been presented. The various systems contrived are portrayed in Table IV and their PSNR versus channel SNR performances are characterized by Figs. 8–11. The QT codecs contrived have a programmable bit rate, which can vary over a wide range, as was demonstrated by Fig. 6, and hence lend themselves to a variety of wireless multimedia applications. Due to this video quality/bit rate flexibility, these video codecs are ideal for employment in the intelligent wireless adaptive multimode terminal of the near future, which is the subject of intensive research in the wireless communications community. Before these multimode terminals become a practical reality, the proposed codec can be used for wireless videophony over conventional fixed-rate mobile radio speech channels, such as the Pan-European GSM system, the IS-54 or IS-95 systems, as well as the Japanese digital cellular system. A further potential application can be found in surveillance systems. Our future work is targeted at improving the complexity, video quality, and robustness of the video codec and the overall system performance by contriving adaptive system reconfiguration algorithms.

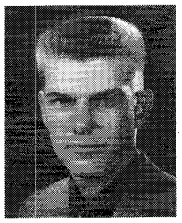
Acknowledgment

The constructive criticism of the anonymous reviewers is gratefully acknowledged.

References

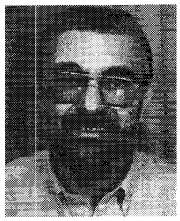
[1] *Proc. of the 2nd Int. Workshop on Mobile Multimedia Communications*, Bristol, UK, vol. MoMuC-2, Apr. 11–13, 1995.
[2] ISO/IEC 11172 MPEG 1, Int. Standard, Coding of Moving Pictures and Associated Audio for Digital Storage Media Up to About 1.5 Mbit/s, Parts 1–3.
[3] ISO/IEC CD 13818 MPEG 2, Int. Standard, Information Technology, Generic Coding of Moving Video and Associated Audio Information, Parts 1–3.
[4] *IEEE Trans. Circuits Syst. Video Technol.*, Special Issue on "Very low bit rate video coding," vol. 4, no. 3, pp. 213–357, June 1994.
[5] *IEEE J. Select. Areas Commun.*, Special Issue on Speech and Image Coding, vol. 10, no. 5, pp. 793–976, June 1992.
[6] J. W. Woods Ed., *Subband Image Coding*. Amsterdam, The Netherlands, Kluwer Academic, 1991.
[7] R. Stedman, H. Gharavi, L. Hanzo, and R. Steele, "Transmission of subband-coded images via mobile channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 1, pp. 15–27, Feb. 1993.
[8] J. Katto, J. Ohki, S. Nogaki, and M. Ohta, "A wavelet codec with overlapped motion compensation for very low bit-rate environment," *IEEE Trans. Video Technol.*, vol. 4, no. 3, pp. 328–338, June 1994.
[9] P. Strobach, "Tree-structured scene adaptive coder," *IEEE Trans. Commun.*, vol. 38, pp. 477–486, 1990.

[10] J. Vaisey and A. Gersho, "Image compression with variable block size segmentation," *IEEE Trans. Signal Processing*, vol. 40, no. 8, pp. 2040–2060, Aug. 1992.
[11] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Amsterdam, The Netherlands: Kluwer Academic, 1992.
[12] B. Ramamurthi and A. Gersho, "Classified vector quantization of images," *IEEE Trans. Commun.*, vol. 31, no. 11, pp. 1105–1115, Nov. 1986.
[13] J. F. Arnold, X. Zhang, and M. C. Cavenor, "Adaptive quadtree coding of motion-compensated image sequences for use on the broadband ISDN," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 222–229, June 1993.
[14] E. Shustermann and M. Feder, "Image compression via improved quadtree decomposition algorithms," *IEEE Trans. Image Processing*, vol. 3, no. 2, pp. 207–215, Mar. 1994.
[15] *Proc. of the Int. Workshop on Coding Techniques for Very Low Bit-Rate Video*, Hi-Vision Hall, Shinagawa, Tokyo, Nov. 8–10, 1995.
[16] P. Sheldon, J. Cosmas, and A. Permain, "Dynamically adaptive control system for MPEG-4," in *Proc. of the 2nd Int. Workshop on Mobile Multimedia Communications*, Bristol, UK, Apr. 11–13, 1995.
[17] Advanced Communications Technologies and Services (ACTS), Workplan, DGXIII-B-RA946043-WP, Commission of the European Community, Brussels, 1994.
[18] L. Hanzo and J. Stefanov, "The pan-european digital cellular mobile radio system—known as GSM," in *Mobile Radio Communications*, R. Steele, Ed. London: IEEE Press-Pentech Press, pp. 677–773, ch. 8, 1992.
[19] D. Greenwood and L. Hanzo, "Mobile radio channels," in *Mobile Radio Communications*, R. Steele, Ed. London: IEEE Press-Pentech Press, pp. 92–185, ch. 2, 1992.
[20] K. H. H. Wong and L. Hanzo, "Channel coding," in *Mobile Radio Communications*, R. Steele, Ed. London: IEEE Press-Pentech Press, pp. 347–489, ch. 4, 1992.
[21] W. T. Webb and L. Hanzo, *Modern quadrature amplitude modulation: Principles and applications for fixed and wireless channels.* London: IEEE Press-Pentech Press, p. 557, 1994.
[22] A. K. Jain, *Fundamentals of Digital Image Processing.* London: Prentice-Hall, 1989.
[23] E. Biglieri and M. Luise, "Coded modulation and bandwidth-efficient transmission," in *Proc. of the Fifth Tirrenia Int. Workshop*, 1992, Elsevier, Netherlands, Sept. 8–12, 1991.
[24] L. F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, vol. 41, no. 10, pp. 1439–1450, Oct. 1993.
[25] J. M. Torrance and L. Hanzo, "A comparative study of pilot symbol assisted modem schemes," in *Proc. of RRAS '95*, Bath, UK, Conf. Public., Sept. 26–28, 1995, no. 415, pp. 36–41.
[26] L. Hanzo and J. P. Woodard, "An intelligent multimode voice communications system for indoors communications," *IEEE Trans Veh. Technol.*, vol. 44, no. 4, pp. 735–749, Nov. 1995.
[27] S. Lin, D. J. Costello, and M. J. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.*, pp. 5–17, Dec. 1984.
[28] R. V. Cox, J. Hagenauer, N. Seshadri, and C. E. W. Sundberg, "Subband speech coding and matched convolutional channel coding for mobile radio channels," *IEEE Trans. Signal Processing*, vol. 39, no. 8, pp. 1717–1731, Aug. 1991.



Jürgen Streit was born in Cologne, Germany in 1968. He received the Dipl.-Ing. degree in electronic engineering from the Aachen University of Technology in 1993.

Since 1992 he has been with the Department of Electronics and Computer Science, University of Southampton, UK, working with the mobile multimedia communications research group. He is currently working toward his Ph.D. in image coding.



Lajos Hanzo (M'91–SM'92) graduated in electronics in 1976 and in 1983 he was conferred a Ph.D.

During his 20-year career in telecommunications he has held various research and academic posts in Hungary, Germany, and the UK. Since 1986 he has been with the Department of Electronics and Computer Science, University of Southampton, UK and has been a consultant to Multiple Access Communications Ltd., UK. He co-authored two books on mobile radio communications, published about 100 research papers and was awarded a number of distinctions.

Dr. Hanzo is a member of the IEE.